Quantile Regression Analysis of Length-Biased Survival Data

Huixia Wang and Lan Wang

05/18/2011

Abstract

Analysis of length-biased time-to-event data, which commonly arise in epidemiological cohort studies and cross-sectional surveys, has attracted considerable attention recently. Ignoring length-biased sampling often leads to severe bias in estimating the survival time in the general population. Existing work either completely ignore the covariate effects or use hazard or accelerated failure time regressions, which restrict the covariates to affect only the location of the transformed survival distribution. In this paper, we propose a flexible quantile regression framework for analyzing the covariate effects on the population survival time under both length-biased sampling and random right censoring. This framework allows for easy interpretation of the statistical model. Furthermore, it allows the covariates to have different impacts at different tails of the survival distribution and thus is able to capture important population heterogeneity. Using an unbiased estimating equation approach, we develop two estimators, one for covariate-independent censoring and the other for covariate-dependent censoring. We establish the consistency and asymptotic normality theory for both estimators. A lack-of-fit test is proposed for diagnosing the adequacy of the population quantile regression model. The finite sample performance of the proposed methods is assessed through a simulation study. By analyzing the CSHA dementia and the Spain unemployment data sets, we demonstrate that the proposed methods are effective in discovering interesting covariate effects at different tails of the target distribution, which were largely overlooked by existing methods in the literature.

Key Words: censored quantile regression; length-biased data; Kaplan-Meier estimator; kernel; survival time.

[1]Huixia Wang is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695. Lan Wang is Associate Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455. The research is supported by NSF grants DMS-0706963, DMS-1007420 and DMS-1007603. H. Wang and L. Wang contributed equally to this work.

1 Introduction

In biomedical studies, economics and many other areas, we are often interested in studying the time that elapses before an event is experienced, for example, the time from the onset of a disease to the death of a patient or the time from losing a job to reemployment. In observational studies, such time-to-event data are often complicated by both right censoring and length-biased sampling. Right censoring occurs when the event of interest (e.g., death of a patient, reemployment) has not been observed at the end of the study; while length-biased sampling occurs when the event has already occurred before the recruitment time of the study. For example, the Canadian Study of Health and Aging (CSHA, Asgharian et al., 2002; Shen et al., 2009) data set we analyzed in Section 4.1 contains information on over eight hundred Canadian seniors over the age of 65 who were screened for dementia in a study starting in 1991. About 22% of the participants were still alive (thus censored) when the study ended. On the other hand, those seniors who had dementia but died before the sampling time (or recruitment time) were not included in the study. Since the seniors with dementia had to survival long enough to be included in the sample, the observed sample tends to overestimate the survival time of the general population. The Spain unemployment data we analyzed in Section 4.2 display similar features. In this paper we use length-biased sampling to refer to the sampling scheme under which the probability of selecting an individual in the population is proportional to the length of the target time variable.

Ignoring length-biased sampling may lead to severe bias in estimation. Considerable work have been developed to correct the bias due to length-biased sampling. Existing literature on this problem can be roughly divided into two classes. One class of work uses the unconditional approach, which estimates the marginal survival distribution in the general population without using the covariates information. Wang (1991), de Uña Álvarez (2002), Asgharian et al. (2002), Luo & Tsai (2009) and Huang & Qin (2011) proposed nonparametric methods for estimating the marginal survival distribution. For studies with two left-truncated time-to-event variables, Jiang et al. (2005) proposed a semiparametric estimation of the marginal distribution of the nonterminal event. Ning et al. (2010) developed a nonparametric test procedure to assess the equality of two marginal survival functions.

Another class of work developed various semiparametric regression models to account for the covariate effects. For length-biased data without right censoring, Wang (1996) developed a semiparametric estimation method for the Cox Proportional Hazard (PH) model, and Shin (2009) proposed a two-stage rank estimation procedure for the accelerated failure time (AFT) model. de Uña Álvarez & Iglesias-Pérez (2010) proposed a nonparametric estimation of the conditional distribution of the survival times given the covariates assuming that the censoring times are observed. For length-biased data subject to random right censoring, different estimating procedures were developed by Shen et al. (2009) for the transformation and AFT models, Qin & Shen (2010) for the Cox PH model, Shen (2009) for the proportional and additive hazards models, and Ning et al. (2011) for the AFT models. It is known that these popular semiparametric models such as the PH and AFT models restrict the covariates to affect only the location but not the shape of the distribution of transformed survival times (Koenker & Geling, 2001). Consequently, some important forms of heterogeneity might be overlooked.

In this paper, we propose a flexible quantile regression framework for analyzing time-to-event data that are randomly right censored and length biased. Unlike hazard regression, quantile regression models the conditional quantiles of the survival time directly. Therefore the results are easier to interpret. More importantly, quantile regression allows the covariates to have different effects at different tails of the survival distribution and thus is able to capture important population heterogeneity (Koenker, 2005). Existing work on censored quantile regression include Ying et al. (1995), McKeague et al. (2001), Bang & Tsiatis (2002), Honoré et al. (2002), Portnoy (2003), Peng & Huang (2008), Wang & Wang (2009), among others. However, these work all assume that the sample under analysis is a random sample from the population of interest. They do not apply to censored data from observational studies with length-biased sampling.

Let $\widetilde{T}$ denote the time measured from an *initiation event* (e.g., onset of a disease, loss of a job) to the event of interest (e.g. death, reemployment) in a population of interest. Following the terminology in analysis of time-to-event data in biomedical studies, we refer to $\widetilde{T}$ as the *population survival time*, and refer to the event of interest as *failure*. We are interested in estimating the conditional quantiles of $\widetilde{T}$ given the observed covariates $\mathbf{x}$.

We derive unbiased estimating equations for estimating the parameters in the conditional quantile regression model. Two different estimators are proposed: one assumes that the censoring time is independent of the covariates (a common assumption in the literature) and the other allows the censoring time to depend on the covariates. We establish consistency and asymptotic normality theory for both estimators. Furthermore, we propose a lack-of-fit test for model checking and establish relevant asymptotic theory.

The rest of the paper is organized as follows. In Section 2, we present two estimation methods together with their asymptotic properties, and a lack-of-fit test for model checking. In Section 3, we assess the finite sample performance of the proposed estimation and lack-of-fit test through a simulation study. In Section 4, we illustrate the value of the proposed methods by analyzing the CSHA dementia and Spain unemployment data sets. We demonstrate that the proposed quantile regression methods not only produce results with easy interpretation, but also help discover interesting covariate effects at different tails of the target distribution, which were overlooked when applying existing methods. All the technical details are given in the Appendix.

2 Proposed Methods

2.1 Notations and Quantile Regression Model

Following the notation in Shen et al. (2009), we use A to denote the time of sampling measured from the initiation event. In the literature, A is also referred to as the *truncation variable* or *backward recurrence time*. We use V to denote the time measured from the sampling time to failure (*residual survival time*). Suppose that V is randomly censored by a random variable C, which is called the *residual potential censoring variable* in Asgharian et al. (2002). Let $T = A + V$ be the observed survival time. One can only observe T if the population survival time $\widetilde{T} > A$. We assume that C and (A, V) are independent conditional on the covariates $\mathbf{x}$.

Let $f_T(t|\mathbf{x})$ be the conditional density function of T given $\mathbf{x}$ and let $f_{\widetilde{T}}(t|\mathbf{x})$ be the conditional density function of $\widetilde{T}$ given $\mathbf{x}$. As in Wang (1991), Asgharian et al. (2002), Shen et al. (2009), among others, the length-biased data are assumed to satisfy the stationarity assumption, i.e., the initiation times follow a stationary Poisson process. Under the stationarity condition, which is reasonable for many real applications, we have

$$f_T(t|\mathbf{x}) = \frac{t f_{\widetilde{T}}(t|\mathbf{x})}{\mu(\mathbf{x})}, \tag{1}$$

where $\mu(\mathbf{x}) = \int_0^\infty t f_{\widetilde{T}}(t|\mathbf{x}) dt$. It can also be derived that (Zelen, 2005; Asgharian & Wolfson, 2005) the joint conditional density function of (A, V) given $\mathbf{x}$ (also the conditional density function of (A, T)) is

$$f_{A,V}(a, v|\mathbf{x}) = f_{\widetilde{T}}(a + v|\mathbf{x}) I(a > 0, v > 0)/\mu(\mathbf{x}). \tag{2}$$

In the sequel, we let $F_0(\cdot|\mathbf{x})$ denote the conditional distribution function of V and let $G_0(s|\mathbf{x})$ denote the conditional survival function of C, i.e., $F_0(s|\mathbf{x}) = P(V \le s|\mathbf{x})$ and $G_0(s|\mathbf{x}) = P(C > s|\mathbf{x})$. Let $w_0(t|\mathbf{x}) = \int_0^t G_0(s|\mathbf{x}) ds$. In addition, let $f_C(s|\mathbf{x})$ be the conditional density function of C given $\mathbf{x}$.

We observe a random sample $(Y_i, \mathbf{x}_i, \delta_i)$, $i = 1, \ldots, n$, where

$$Y_i = \min(T_i, A_i + C_i), \quad T_i = A_i + V_i \quad \text{and} \quad \delta_i = I(V_i \le C_i),$$

and $\mathbf{x}_i$ is the p-dimensional vector of covariates. We also observe the initiation time A_i. Let $\tau \in (0, 1)$ be a given quantile level. To estimate the τth conditional quantile of $\widetilde{T}_i$, we assume that $\widetilde{T}_i$ (or a transformation of $\widetilde{T}_i$ such as the log transformation) follows the quantile regression model

$$\widetilde{T}_i = \mathbf{z}_i^T \boldsymbol{\beta}_0(\tau) + \epsilon_i, \tag{3}$$

where $\mathbf{z}_i = (1, \mathbf{x}_i^T)^T$, $\boldsymbol{\beta}_0(\tau) = (\alpha_0(\tau), \boldsymbol{\gamma}_0^T(\tau))^T$, and the random error ϵ_i satisfies $P(\epsilon_i \le 0|\mathbf{x}_i) = \tau$. For $\tau = 1/2$, $\mathbf{z}_i^T \boldsymbol{\beta}_0(1/2)$ is the conditional median of the population survival time. The random errors $\epsilon_1, \ldots, \epsilon_n$ are independent, but we do not require them to be identically distributed. For notational simplicity, in the sequel we ignore τ in $\beta_0(\tau)$ when confusion is not made.

2.2 Unbiased Estimating Equation

To motivate the estimating equation for $\boldsymbol{\beta}_0$ in (3), we first consider the ideal case where $G_0(s|\mathbf{x})$, the conditional survival function of C, is known. Let E denote the expectation taken under the true underlying distribution of the data. We have

$$
\begin{aligned}
& E\Big[\frac{\delta_i\{I(Y_i \le \mathbf{z}_i^T\boldsymbol{\beta}) - \tau\}}{w_0(Y_i|\mathbf{x}_i)}\Big|\mathbf{x}_i = \mathbf{x}\Big] \\
= \; & E\Big[\frac{I(V_i \le C_i)\{I(T_i \le \mathbf{z}_i^T\boldsymbol{\beta}) - \tau\}}{w_0(Y_i|\mathbf{x}_i)}\Big|\mathbf{x}_i = \mathbf{x}\Big] \\
= \; & E\Big[\frac{I(V_i \le C_i)\{I(A_i + V_i \le \mathbf{z}_i^T\boldsymbol{\beta}) - \tau\}}{w_0(A_i + V_i|\mathbf{x}_i)}\Big|\mathbf{x}_i = \mathbf{x}\Big] \\
= \; & \iiint \frac{I(v < c)}{w_0(a + v|\mathbf{x}_i)}\{I(a + v \le \mathbf{z}^T\boldsymbol{\beta}) - \tau\} f_{A,V}(a, v|\mathbf{x}) f_C(c|\mathbf{x})\,\mathrm{d}c\,\mathrm{d}a\,\mathrm{d}v \\
= \; & \iint \frac{G_0(v|\mathbf{x})}{w_0(a + v|\mathbf{x})}\{I(a + v \le \mathbf{z}^T\boldsymbol{\beta}) - \tau\} f_{A,V}(a, v|\mathbf{x})\,\mathrm{d}a\,\mathrm{d}v \\
= \; & \int_0^\infty \int_0^y \frac{G_0(v|\mathbf{x})}{w_0(y|\mathbf{x})}\{I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau\} f_{\widetilde{T}}(y|\mathbf{x}) I(y - v > 0, v > 0)/\mu(\mathbf{x})\,\mathrm{d}v\,\mathrm{d}y \\
= \; & \frac{1}{\mu(\mathbf{x})}\int_0^\infty \{I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau\} f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y,
\end{aligned}
$$

where the third equality uses the conditional independence between C and (A, V), the second last equality applies the transformation $y = a + v$ and formula (2).

Note that the above expectation is zero when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Thus, when $G_0(s|\mathbf{x})$ is known, an unbiased estimating equation for $\boldsymbol{\beta}_0$ is

$$\sum_{i=1}^n \delta_i \mathbf{z}_i \frac{\{I(Y_i \le \mathbf{z}_i^T\boldsymbol{\beta}) - \tau\}}{w_0(Y_i|\mathbf{x}_i)} = 0.$$

2.3 Proposed Estimation Methods

In practice, $G_0(s|\mathbf{x})$ and hence $w_0(t|\mathbf{x})$ is unknown. In the following, we consider two feasible estimating equations under different assumptions of the censoring variable distribution.

Under the assumption that the censoring variable C_i and the covariates $\mathbf{x}_i$ are independent, we may estimate $G_0(s|\mathbf{x})$ by the classical (or global) Kaplan-Meier estimator, defined as $\widehat{G}_1(s)$, which is obtained from $(Y_j - A_j, \delta_j)$, $j = 1, \ldots, n$. In this case, we replace $w_0(Y_i|\mathbf{x})$

with $\int_0^{Y_i} \widehat{G}_1(s)\mathrm{d}s$ and estimate $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}_1$ that solves the following estimating equation

$$\mathbf{M}_n(\boldsymbol{\beta}, \widehat{G}_1) = n^{-1} \sum_{i=1}^{n} \delta_i \mathbf{z}_i \frac{\{I(Y_i \leq \mathbf{z}_i^T \boldsymbol{\beta}) - \tau\}}{\int_0^{Y_i} \widehat{G}_1(s)\,\mathrm{d}s} = 0. \tag{4}$$

Alternatively, we consider a more relaxed assumption where the distribution of C_i may depend on the covariates $\mathbf{x}_i$. We adopt the following local Kaplan-Meier estimator

$$\widehat{G}_2 = \widehat{G}_2(s|\mathbf{x}) \;=\; \prod_{j=1}^{n} \left\{ 1 - \frac{B_{nj}(\mathbf{x})}{\sum_{k=1}^{n} I(Y_k - A_k \geq Y_j - A_j) B_{nk}(\mathbf{x})} \right\}^{I(Y_j - A_j \leq s, \delta_j = 0)}, \tag{5}$$

where $\{B_{nk}(\mathbf{x}), k = 1, \cdots, n\}$ is a sequence of nonnegative weights adding up to 1. When $B_{nj}(\mathbf{x}) = 1/n$ for all j, $\widehat{G}_2 = \widehat{G}_1$. Here, we employ the Nadaraya-Watson weights which are defined as $B_{nk}(\mathbf{x}) = K(\frac{\mathbf{x}-\mathbf{x}_k}{h_n})/\{\sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{x}_i}{h_n})\}, k = 1, \ldots, n$. In the above definition, $h_n \in \mathbb{R}^+$ is the bandwidth, $K(\frac{\mathbf{x}-\mathbf{x}_i}{h_n}) = K(\frac{x_1 - x_{i1}}{h_n}, \ldots, \frac{x_p - x_{ip}}{h_n})$, where x_{ij} is the jth element of $\mathbf{x}_i$. We adopt the commonly used product kernel function $K(u_1, \ldots, u_p) = \prod_{i=1}^{p} K(u_i)$ with $K(\cdot)$ being a univariate kernel function. We replace $w_0(Y_i|\mathbf{x})$ with $\int_0^{Y_i} \widehat{G}_2(s|\mathbf{x})\mathrm{d}s$, where $\widehat{G}_2(s|\mathbf{x})$ is defined in (5). We then estimate $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}_2$ that solves the following estimating equation

$$\mathbf{M}_n(\boldsymbol{\beta}, \widehat{G}_2) = n^{-1} \sum_{i=1}^{n} \delta_i \mathbf{z}_i \frac{\{I(Y_i \leq \mathbf{z}_i^T \boldsymbol{\beta}) - \tau\}}{\int_0^{Y_i} \widehat{G}_2(s|\mathbf{x}_i)\,\mathrm{d}s} = 0. \tag{6}$$

In the following, we define

$$\begin{aligned} Q_n(\boldsymbol{\beta}, \widehat{G}_1) &= n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\int_0^{Y_i} \widehat{G}_1(s)\,\mathrm{d}s} \rho_\tau(Y_i - \mathbf{z}_i^T \boldsymbol{\beta}), \\ Q_n(\boldsymbol{\beta}, \widehat{G}_2) &= n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\int_0^{Y_i} \widehat{G}_2(s|\mathbf{x}_i)\,\mathrm{d}s} \rho_\tau(Y_i - \mathbf{z}_i^T \boldsymbol{\beta}), \end{aligned}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function. Then it is easy to see that $\mathbf{M}_n(\boldsymbol{\beta}, \widehat{G}_i)$ is the negative gradient function of $Q_n(\boldsymbol{\beta}, \widehat{G}_i)$, $i = 1, 2$. Since the estimating equations (4) and (6) are discontinuous, an exact zero solution may not exist. To avoid ambiguity and facilitate computation, we formally define $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ as the minimizer of $Q_n(\boldsymbol{\beta}, \widehat{G}_1)$ and $Q_n(\boldsymbol{\beta}, \widehat{G}_2)$, respectively. As $Q_n(\boldsymbol{\beta}, \widehat{G}_i)$ is convex in $\boldsymbol{\beta}$, $i = 1, 2$, it can be easily minimized using linear programming and existing statistical software packages.

2.4 Asymptotic Properties

The two estimators $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ defined in Section 2.3 both approximately solve semiparametric estimating equations. Different from classical M-estimators, the estimating equations here are discrete and depend on a preliminary nonparametric estimator of the unknown function $G_0(s|\mathbf{x})$. To establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}_i$, $i = 1, 2$, we utilize the theory of Chen et al. (2003) for non-smooth estimating equations that may depend on an infinite-dimensional nuisance parameter. This theory was developed using modern empirical processes techniques.

We impose the following regularity conditions.

(C1) The parameter space $\mathbb{B}$ is a compact subset of $\mathbb{R}^{p+1}$ and the true parameter value $\boldsymbol{\beta}_0$ is in the interior of $\mathbb{B}$.

(C2) The covariate vector $\mathbf{x}$ has a joint density function $h(\mathbf{x})$ whose νth derivative is bounded from zero and infinity uniformly in $\mathbf{x}$, $\nu \geq 2$. Furthermore, $P(||\mathbf{x}|| \leq L) = 1$ for a finite positive constant L.

(C3) The conditional survival function of C satisfies: $t_0 \leq \sup\{t : G_0(t|\mathbf{x}) > 0\} \leq t_1$ for all $\mathbf{x}$ and some finite positive constants t_0 and t_1.

(C4) The functions $F_0(s|\mathbf{x})$ and $G_0(s|\mathbf{x})$ have continuous first derivatives with respect to s for each $\mathbf{x}$, and they have bounded (uniformly in s) νth order partial derivatives with respect to $\mathbf{x}$, where $\nu \geq 2$ is a positive integer.

(C5) The random error ϵ has a continuous density $f_\epsilon(s|\mathbf{x})$ that is uniformly bounded away from zero for $0 \leq s \leq K$ and $\mathbf{x} \in \mathbb{X} = \{\mathbf{x} \in \mathbb{R}^p : ||\mathbf{x}|| \leq L\}$, where K is a finite positive constant and L is the constant in condition (C2).

(C6) For any $\mathbf{x}$, $E\{Y \min(Y, t_0)^{-4}|\mathbf{x}\} \leq C$, where C is a positive constant.

(C7) $\lambda_{min}\big[\mathrm{E}\{\mathbf{x}\mathbf{x}^T/\mu(\mathbf{x})\}\big] \geq \eta$, where η is a finite positive constant, and $\lambda_{min}(\mathbf{A})$ denotes the smallest eigenvalue of matrix $\mathbf{A}$.

(C8) The univariate nonnegative kernel function $K(\cdot)$ has a compact support. It is a νth order kernel function satisfying $\int K(u)du = 1$, $\int K^2(u)du < \infty$, $\int u^j K(u)du = 0$ for $j < \nu$ and $\int |u|^\nu K(u)du < \infty$, and it is Lipschitz continuous of order ν, where $\nu \geq 2$ is the positive integer in condition (C4).

We write $\mathbf{M}_n(\boldsymbol{\beta}, G) = n^{-1}\sum_{i=1}^n m(\mathbf{W}_i, \boldsymbol{\beta}, G)$, where $\mathbf{W}_i = (\mathbf{x}_i, Y_i, \delta_i)^T$ denotes the data from the ith observation and $m(\mathbf{W}_i, \boldsymbol{\beta}, G) = \delta_i \mathbf{z}_i \{I(Y_i \leq \mathbf{z}_i^T \boldsymbol{\beta}) - \tau\}\{\int_0^{Y_i} G(s|\mathbf{x}_i)\, \mathrm{d}s\}^{-1}$. As the sample size n gets large, we expect that $\mathbf{M}_n(\boldsymbol{\beta}, G)$ converges in probability to the population version estimating function

$$\mathbf{M}(\boldsymbol{\beta}, G) = E[m(\mathbf{W}_i, \boldsymbol{\beta}, G)] = E\Big[\frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty \frac{\int_0^y G_0(s|\mathbf{x})\, \mathrm{d}s}{\int_0^y G(s|\mathbf{x})\, \mathrm{d}s} [I(y \leq \mathbf{z}^T \boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\, \mathrm{d}y\Big]$$

following similar calculations as in Section 2.2. Note that $\mathbf{M}(\boldsymbol{\beta}_0, G_0) = 0$. Therefore, $\mathbf{M}_n(\boldsymbol{\beta}, G)$ satisfies the conditions in Theorems 2 and 3 in Chen et al. (2003). The following theorems summarize the consistency and asymptotic normality properties of the two estimators $\widehat{\boldsymbol{\beta}}_i$, $i = 1, 2$.

Theorem 2.1 *(Consistency)*
(1) Assume that conditions (C1)-(C7) hold and that the censoring variable C is independent of V and $\mathbf{x}$. Then we have $\widehat{\boldsymbol{\beta}}_1 \to \boldsymbol{\beta}_0$ in probability.
(2) Assume that conditions (C1)-(C8) hold and that the censoring variable C is independent of V given $\mathbf{x}$. If $h_n = n^{-\alpha}$ with $0 < \alpha < \min(1/\nu, 1/p)$, then $\widehat{\boldsymbol{\beta}}_2 \to \boldsymbol{\beta}_0$ in probability.

The asymptotic normality of $\widehat{\boldsymbol{\beta}}_i$ is characterized by how the estimating function $\mathbf{M}(\boldsymbol{\beta}, G)$ depends on the finite-dimensional component $\boldsymbol{\beta}$ and the infinite-dimensional component $\mathbf{G}$. More specifically, we need to characterize how $\mathbf{M}(\boldsymbol{\beta}, G)$ changes as we perturb $\boldsymbol{\beta}$ in a small neighborhood of $\boldsymbol{\beta}_0$ or when we perturb G in the direction of $\widehat{G}_i - G_0$. For this purpose, we consider the ordinary derivative

$$\boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0) = \frac{\partial \mathbf{M}(\boldsymbol{\beta}, G_0)}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} = E\Big\{\frac{\mathbf{z}\mathbf{z}^T}{\mu(\mathbf{x})} f_\epsilon(0|\mathbf{x})\Big\},$$

and the pathwise derivative in the direction $G - G_0$

$$\boldsymbol{\Gamma}_2(\boldsymbol{\beta}, G_0)[G - G_0] = \lim_{a \to 0} \frac{\mathbf{M}\{\boldsymbol{\beta}, G_0 + a(G - G_0)\} - \mathbf{M}(\boldsymbol{\beta}, G_0)}{a}.$$

The asymptotic distributions of $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ rely on $\boldsymbol{\Gamma}_2(\boldsymbol{\beta}, G_0)[\widehat{G}_i - G_0]$ and differer in that $\widehat{G}_i - G_0$, $i = 1, 2$, have different linear representations of independent and identically distributed (i.i.d.) random variables. For the classical Kaplan-Meier estimator, an almost sure i.i.d. representation of $\widehat{G}_1(s) - G_0(s)$ is given in Lo & Singh (1985); while for the local Kaplan-Meier estimator, an almost sure i.i.d. representation of $\widehat{G}_2(s|\mathbf{x}) - G_0(s|\mathbf{x})$ can be found in

Gonzalez-Manteiga & Cadarso-Suarez (1994).

Theorem 2.2 *(Asymptotic normality)*
(1) Assume that conditions (C1)-(C7) hold and that the censoring variable C is independent of V and $\mathbf{x}$, then

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0)^{-1}\mathbf{V}_1\boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0)^{-1}).$$

where $\mathbf{V}_1 = Cov\{m(\mathbf{W}_i, \boldsymbol{\beta}_0, G_0) + \phi_1(V_j^*, \delta_j, \mathbf{x}_j)\}$ with ϕ_1 and V_j^* defined in the appendix.
(2) Assume that conditions (C1)-(C8) hold, the censoring variable C is independent of V given the covariates $\mathbf{x}$, $h_n = n^{-\alpha}$ for some $\frac{1}{2\nu} < \alpha < \frac{1}{3p}$, where $\nu > 3/2p$, then

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0)^{-1}\mathbf{V}_2\boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0)^{-1}),$$

where $\mathbf{V}_2 = Cov\{m(\mathbf{W}_i, \boldsymbol{\beta}_0, G_0) + \phi_2(V_j^*, \delta_j, \mathbf{x}_j)\}$ with ϕ_2 and V_j^* defined in the appendix.

Remark. It is seen from Theorem 2.2 that the asymptotic covariance matrices of both $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ depend on the unknown density function $f_\epsilon(0|\mathbf{x})$, which is difficult to estimate in practice. For variance estimation, we adopt a simple bootstrap procedure by resampling the quadruples $(Y_i, \mathbf{x}_i, \delta_i, A_i)$ with replacement. The covariance matrix of the proposed estimator $\widehat{\boldsymbol{\beta}}_i, i = 1, 2$, is then estimated by the sample covariance matrix of the analogous bootstrap estimates.

The proposed estimator $\widehat{\boldsymbol{\beta}}_2$ involves the local Kaplan-Meier estimates and thus requires choosing the bandwidth parameter h_n. Our empirical study in Section 3.3 suggests that the performance of the proposed procedure is not overly sensitive to the choice of h_n. For practical data analysis, we suggest to use the computationally more intensive K-fold cross validation to choose the bandwidth. We first divide the data set randomly into K parts with roughly equal size. For the kth part, $k = 1, \cdots, K$, we fit model (3) using the rest $K-1$ parts of the data, and then evaluate the quantile loss from predicting the τth conditional quantile of $\widetilde{T}$. Here the quantile loss is calculated by $\sum_{i\in\mathcal{I}_k} \delta_i\{\int_0^{Y_i} \widehat{G}_2(s|\mathbf{x}_i)\,\mathrm{d}s\}^{-1}\rho_\tau\{Y_i - \mathbf{z}_i^T\widehat{\boldsymbol{\beta}}_{(-k)}\}$, where $\mathcal{I}_k$ contains the indices of cases belonging to the kth part and the $\widehat{\boldsymbol{\beta}}_{(-k)}$ is the local quantile coefficient estimate with the kth part data excluded. We choose the h_n that gives the minimum average quantile loss.

2.5 Model Diagnosis

Model diagnosis is important in real data analysis. To check the validity of the linear model assumption in (3), we consider $e_i = \delta_i\{I(Y_i \leq \mathbf{z}_i^T\boldsymbol{\beta}_0) - \tau\}w_0^{-1}(Y_i|\mathbf{x}_i)$, $i = 1, \dots, n$. Calculations in Section 2.2 indicate that $E(e_i|\mathbf{x}_i) = 0$ when the model in (3) is correctly specified. We may therefore construct a statistical test based on the pseudo residuals $\widehat{e}_i = \delta_i\{I(Y_i \leq \mathbf{z}_i^T\widehat{\boldsymbol{\beta}}) - \tau\}\widehat{w}^{-1}(Y_i|\mathbf{x}_i)$, $i = 1, \dots, n$, where $\widehat{\boldsymbol{\beta}}$ and $\widehat{w}^{-1}(Y_i|\mathbf{x}_i)$ are appropriate estimators of $\boldsymbol{\beta}_0$ and $w_0(Y_i|\mathbf{x}_i)$, respectively.

More specifically, we consider the following conditional moment test statistic

$$T_n = \frac{1}{n(n-1)\lambda_n^p}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\lambda_n}\right)\widehat{e}_i\widehat{e}_j, \tag{7}$$

where λ_n is a smoothing parameter, and $K(\cdot)$ is the product kernel function; see the discussion after (6).

The statistic T_n provides a nonparametric kernel estimator of $E\{E(e_i|\mathbf{x}_i)^2 h(\mathbf{x}_i)\}$, which is zero if and only if (3) is correctly specified. Conditional moment test for quantile regression with uncensored data was considered by Zheng (1998) and was extended to the random censoring case in Wang (2008). However, these tests are not applicable to length biased survival data.

Theorem 2.3 *(1) Assume that the conditions of Theorem 2.2(1) are satisfied. Let T_{n1} be the test statistic in (7) with $\widehat{e}_i$ defined as $\delta_i\{I(Y_i \leq \mathbf{z}_i^T\widehat{\boldsymbol{\beta}}_1) - \tau\}\{\int_0^{Y_i}\widehat{G}_1(s)\}^{-1}$. If model (3) is correctly specified, $\lambda_n \to 0$ and $n^\gamma\lambda_n^{2p} \to \infty$ for some $0 < \gamma < 1$, then $n\lambda_n^{p/2}T_{n1} \to N(0, \kappa_1^2)$, where $\kappa_1^2 = E[\delta_i\{I(Y_i \leq \mathbf{z}_i^T\boldsymbol{\beta}_0) - \tau\}\{\int_0^{Y_i} G_0(s)ds\}^{-1}]^2$.*
(2) Assume that the conditions of Theorem 2.2(2) are satisfied. Let T_{n2} be the test statistic in (7) with $\widehat{e}_i$ defined as $\delta_i\{I(Y_i \leq \mathbf{z}_i^T\widehat{\boldsymbol{\beta}}_2) - \tau\}\{\int_0^{Y_i}\widehat{G}_2(s|\mathbf{x}_i)\}^{-1}$. If model (3) is correctly specified, $\lambda_n \to 0$ and $n^\gamma\lambda_n^{2p} \to \infty$ for some $0 < \gamma < 1$, then $n\lambda_n^{p/2}T_{n2} \to N(0, \kappa_2^2)$, where $\kappa_2^2 = E[\delta_i\{I(Y_i \leq \mathbf{z}_i^T\boldsymbol{\beta}_0) - \tau\}\{\int_0^{Y_i} G_0(s|\mathbf{x}_i)ds\}^2]^2$.

3 Simulation Study

3.1 Simulation Design

We next assess the performance of the proposed estimation and lack-of-fit test methods through simulations.. We generate the survival times $\tilde{T}_i$, $i = 1, \ldots, n$, from the model:

$$\log \widetilde{T}_i = \beta_{(0)} + x_{1i}\beta_{(1)} + x_{2i}\beta_{(2)} + (1 + \gamma x_{1i})\epsilon_i, \tag{8}$$

where $\beta_{(0)} = \beta_{(1)} = \beta_{(2)} = 1$, $x_{1i} \sim \text{Bernoulli}(0.5)$, $x_{2i} \sim U(-0.5, 0.5)$, γ controls the level of heteroscedasticity, and ϵ_i is the random error. The initiation times A_i are generated from the $U(0, u_A)$ distribution, where $u_A > 0$ is a constant. We keep the pairs with $\widetilde{T}_i > A_i$, resulting in the length-biased survival time $T_i = A_i + V_i$. Due to random right censoring, only $Y_i = \min(T_i, A_i + C_i)$ is observed, where C_i is the censoring variable. Under model (8), the τth conditional quantile of $\log \widetilde{T}$ is

$$Q_{\log \widetilde{T}}(\tau | x_{1i}, x_{2i}) = \beta_{(0)}(\tau) + x_{1i}\beta_{(1)}(\tau) + x_{2i}\beta_{(2)}(\tau),$$

where $\beta_{(0)}(\tau) = \beta_{(0)} + Q(\tau)$, $\beta_{(1)}(\tau) = \beta_{(1)} + \gamma Q(\tau)$, $\beta_{(2)}(\tau) = \beta_{(2)}$ and $Q(\tau)$ is the τth quantile of ϵ_i. Therefore, for $\gamma \neq 0$, the covariate x_{1i} exhibits different effects at different quantiles.

We consider three different cases. Case 1 corresponds to a homoscedastic model with normal errors, $\gamma = 0$, $\epsilon_i \sim N(0, 0.5^2)$ and $u_A = 50$. Case 2 has heteroscedastic uniform errors with $\gamma = 1$, $\epsilon_i \sim U(-0.5, 0.5)$, $u_A = 30$. Case 3 has heteroscedastic normal errors with $\gamma = 1$, $\epsilon_i \sim N(0, 0.5^2)$, $u_A = 300$. In Cases 1-2, C_i are generated from $U(0, u_C)$ and thus are independent of the survival times. In Case 3, C_i are generated from the mixture distribution, where $C_i \sim \text{Exp}(\lambda)$ for $x_{i2} < 0$, and $C_i \sim \text{Exp}(0.1\lambda)$ for $x_{i2} > 0$. The u_C and λ are chosen to yield two censoring proportions: 20% and 40%. For each case, the simulation is run 500 times for $n = 200$ and 400 at two quantile levels $\tau = 0.25$ and 0.5.

3.2 Estimation

We compare the following three estimation methods: the naive method, the IPW method and the LIPW method. The naive method ignores the length-biased sampling but accounts for random right censoring. It estimates the coefficients by minimizing

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}_1(Y_i - A_i)} \rho_\tau \big\{ \log(Y_i - A_i) - \beta_{(0)}(\tau) - x_{i1}\beta_{(1)}(\tau) - x_{i2}\beta_{(2)}(\tau) \big\}. \tag{9}$$

The IPW method (the name is due to the connection to the inverse probability weighting method in the missing data literature) gives the estimator $\widehat{\boldsymbol{\beta}}_1(\tau)$, and the LIPW (local inverse probability weighting) method gives the estimator $\widehat{\boldsymbol{\beta}}_2(\tau)$, both defined in Section 2.2.

Tables 1–3 summarize the bias, mean squared error and empirical coverage probabilities of 95% bootstrap confidence intervals for each of the aforementioned three methods in Cases 1–3. The bootstrap confidence intervals are based on 500 bootstrap samples. The estimators of the naive approach substantially overestimate the conditional quantiles of the population survival time $\widetilde{T}$ in all the scenarios considered. In addition, the associated inference of the naive method is inaccurate. In Cases 1–2 with covariate-independent censoring, the IPW method and the LIPW method perform equally well: both lead to essentially unbiased estimates with coverage probabilities close to the nominal level. However, when the censoring depends on the covariate x_{2i} in Case 3, the IPW method based on the global Kaplan-Meier estimate of the censoring distribution gives biased estimates especially for the effect of x_{2i} with 40% censoring, and consequently the associated confidence intervals have coverage probabilities significantly lower than 95%. In summary, the LIPW method is more flexible, and it performs competitively well in both scenarios no matter the censoring distribution depends on the covariates or not.

[Tables 1–3 are about here]

3.3 Sensitivity of LIPW to Bandwidth h_n and Cross-validation

In the above simulation, the bandwidth h_n is chosen as $h_n = n^{-0.3}$ for the LIPW method. To investigate the sensitivity of the LIPW method to h_n, we compare the mean squared error

(MSE) of LIPW estimators with $h_n = dn^{-0.3}$, where $d = 0.1, 0.2, \cdots, 2.0$. Figure 1 plots the mean squared error of the LIPW estimates versus d in Case 2 with $\tau = 0.25$. We observe that the performance of the LIPW method is not sensitive to the choice of h_n. For each curve, the solid point on the right end corresponds to the LIPW estimate with h_n selected by the five-fold cross validation (CV) method as described in Section 2.4. It seems that the MSE of the LIPW method varies very little for $d \in [0.75, 2]$ and the cross validation leads to a good choice of h_n.

[Figure 1 is about here]

3.4 Model Diagnosis

To assess the power of the proposed model diagnosis method, we consider testing the linearity of the conditional median function of $\log \tilde{T}$. The simulation setup is the same as that in Case 2 with 20% censoring except that

$$\log \widetilde{T}_i = \beta_{(0)} + x_{1i}\beta_{(1)} + x_{2i}\beta_{(2)} + a\cos(10x_{2i}) + (1 + x_{1i})\epsilon_i, \tag{10}$$

where a measures the deviation from the linear model.

To investigate the effect of bandwidth λ_n involved in the test, we consider $\lambda_n = cn^{-0.3}$ for $c = 0.5$, 0.8, 1.0, 1.3 and 1.5. Figure 2 plots the estimated power curves of the tests T_{n1} and T_{n2} for $n = 200$, different values of a, and different choices of c. The tests for $c = 0.5$ have slightly inflated Type I errors (0.092 for T_{n1} and 0.088 for T_{n2}), and those for $c = 1.5$ are a bit conservative. However, both tests have levels close to the nominal level of 0.05 for $c = 1.0$ and 1.3. In addition, for all c values, the power curves gradually approach one as a increases.

[Figure 2 is about here]

4 Real Data Analysis

4.1 Analysis of the CSHA Data Set

The Canadian Study of Health and Aging (CSHA) is a large cohort study of dementia. The study identified 1132 seniors with dementia between February 1991 and May 1992. These seniors were followed until 1996. Their dates of dementia onset, the dates of enrollment in the study, and the dates of death or last follow-up were recorded. Excluding the cases with missing values, the data set contains 818 Canadian seniors with dementia, among whom 393 of them have probable Alzheimer's disease, 252 have possible Alzheimer's disease, and 173 have vascular dementia. And 638 of them died before the end of the study (22% censoring).

We apply the proposed method to the following model

$$Q_\tau(\log \tilde{T}_i | x_i) = \beta_{(0)}(\tau) + \beta_{(1)}(\tau) x_{1i} + \beta_{(2)}(\tau) x_{2i}, \ i = 1, \cdots, 818, \tag{11}$$

where x_{1i} and x_{2i} are binary variables indicating whether the ith subject has probable Alzheimer's disease or possible Alzheimer's disease. We use vascular dementia as the baseline. Our preliminary analysis indicated that the censoring distribution is not influenced by the binary covariates. Therefore, we focus on the proposed IPW method in the following.

Table 4 summarizes the estimates of the IPW and naive methods at $\tau = 0.1, \cdots, 0.9$. The values in the parentheses are the standard errors estimated by the bootstrap procedure based on 500 bootstrap samples. By ignoring the length-biased sampling, the naive method overestimates the quantiles of the population survival times. Existing methods, for instance the Buckley-James estimator of Ning et al. (2011) did not detect any significant difference among these three categories. However, by regression at different quantiles using the proposed IPW method, we obtain the following observations. Firstly, the seniors with probable Alzheimer's disease tend to have longer survival time than those with vascular dementia, but the effect is not significant at any of the nine quantile levels. Secondly, at upper quantiles $\tau = 0.7, 0.8, 0.9$, the survival time of the seniors with possible Alzheimer's disease is significantly longer than that of the seniors with vascular dementia.

[Table 4 is about here]

Furthermore, we plot the estimated quantiles (in days) of the three categories of diseases based on the IPW method in Figure 3. The shaded area corresponds to the 95% pointwise bootstrap confidence band of the estimated quantiles of the survival of patients with vascular dementia (baseline), that is, $\exp\{\hat{\beta}_{(0)}(\tau)\}$. Figure 3 suggests that the three categories differ little at lower quantiles, but the quantiles of possible Alzheimer's are clearly larger than those of vascular dementia at upper quantiles.

[Figure 3 is about here]

4.2 Analysis of Spain Unemployment Data

The Spain unemployment data set was taken from the Labour Force Survey of the Spanish Institute for Statistics between 1987 and 1997 (de Uña Álvarez et al., 2003). The data set includes 1009 married women who were unemployed at the time of survey. The starting time of unemployment was recorded for each subject. The response variable $\widetilde{T}$ is the total length of unemployment (in month). The event of interest is defined as "finding a job" or "stop searching for a job". The observed sample is length-biased as the sample was restricted only to the subjects who were in the unemployment stock at the inquiry time. The truncation time A is defined as the time between unemployment origin and the inquiry time. At the end of the study (18 months after the first inquiry), 56% of the subjects still searching for jobs were right censored. In this study, since the censoring is due to the end of follow up, the censoring times were known to be 18 months. To demonstrate the proposed methods, we manually generate random censoring times $C_i \in [0, 18]$ as follows. For each censored case, we generate C_i from the truncated exponential distribution $\mathrm{TExp}(1, 18)$ if the subject is below 31 years old (the median age), and from $\mathrm{TExp}(17, 18)$ otherwise. Here the truncated exponential distribution $\mathrm{TExp}(\lambda, 18)$ refers to the conditional distribution of Z given that $Z < 18$ for $Z \sim \mathrm{Exp}(\lambda)$.

We consider the following quadratic quantile regression model:

$$Q_\tau(\log \tilde{T}_i | x_i) = \beta_{(0)}(\tau) + \beta_{(1)}(\tau) x_i + \beta_{(2)}(\tau) x_i^2, \ i = 1, \cdots, 1009, \tag{12}$$

where x_i is the ith subject's age when entering the unemployment stock, which is standardized to have mean zero and variance 1.

To assess if the quadratic quantile regression model (12) is adequate, we apply the lack-of-fit test procedure described in Section 2.5. The bandwidth λ_n is set as $\lambda_n = cn^{-0.3}$ for various c values. Figure 4 plots the p-values of the lack-of-fit tests based T_{n1} and T_{n2} at four quantile levels $\tau = 0.15, 0.25, 0.45$ and 0.5. The results show that the quadratic quantile regression model fits the data reasonably well at lower quantiles $\tau = 0.15$ and 0.25. However, the test based on T_{n1} indicates somewhat lack-of-fit at $\tau = 0.45$, and both test statistics indicate lack-of-fit of the quadratic model at $\tau = 0.5$ for large values of λ_n. This lack-of-fit around the median is possibly due to the high percentage (56%) of right censoring in this data set.

[Figure 4 is about here]

Table 5 summarizes the coefficient estimation of the naive method, the IPW method and the LIPW method at $\tau = 0.15, 0.25, 0.45$ and 0.5. The bandwidth is chosen as $h = n^{-0.3} = 0.13$ for the LIPW method. The standard errors are obtained by bootstrapping with 500 bootstrap samples. The IPW method and the LIPW method both suggest that Age has a significant linear effect at all four quantile levels. The LIPW method gives slightly larger estimates of the quadratic effect than the IPW method. With the LIPW method, Age displays significant quadratic effects at the central quantiles $\tau = 0.45$ and 0.5. As demonstrated in our simulation, the results of the LIPW method are more trustworthy when the censoring time strongly depends on the covariates.

[Table 5 is about here]

In Figure 5, we plot the estimated conditional quantiles of the unemployment duration with respect to age when the three different estimation methods are applied. The naive method is expected to overestimate the conditional quantiles of unemployment duration. By analyzing the same data set with a nonparametric estimation method, de Uña Álvarez & Iglesias-Pérez (2010) found that younger unemployed women get out of the unemployment stock earlier. However, our proposed methods suggest that the age effect is convex, having a negative effect on the duration from age 17 to about 40, but having a positive effect beyond age 40 at quantile levels $\tau = 0.25, 0.45$ and 0.5. At the lower quantile $\tau = 0.15$, the LIPW method suggests that the unemployment duration is the shortest around age 40, while the

IPW estimates show that the unemployment duration decreases with age within the entire region of the study (17 to 70).

[Figure 5 is about here]

Acknowledgement

We thank Professors Gema Álvarez-Llorente and M. Soledad Otero-Giráldez from the University of Vigo for providing the Spain unemployment data set. We thank Professor Masoud Asgharian for providing the dementia data from the CSHA, which were collected as part of the Canadian Study of Health and Aging. The core study was funded by the Seniors Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada Project 6606-3954-MC(S). Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/Pharmaceutical Manufacturers Association of Canada Health Activity Program, NHRDP Project 6603-1417-302(R), Bayer Incorporated, and the British Columbia Health Research Foundation Projects 38 (93-2) and 34 (96-1). The study was coordinated through the University of Ottawa and the Division of Aging and Seniors, Health Canada.

5 Appendix

We establish the asymptotic theory for $\widehat{\boldsymbol{\beta}}_2$ in Theorems 2.1-2.2 and the asymptotic distribution of T_{n2} in Theorem 2.3, as these are the more challenging cases. The approaches also apply to $\widehat{\boldsymbol{\beta}}_1$ and T_{n1} with minor modifications.

Let $\mathbb{F} = \Big\{ m(\mathbf{w}, \boldsymbol{\beta}, G) : \beta \in \mathbb{B}, G \in \mathbb{G} \Big\}$, where $\mathbb{B}$ is defined in condition (C1), $m(\mathbf{w}, \boldsymbol{\beta}, G) = \delta \mathbf{z} \big[I(Y \le \mathbf{z}^T \boldsymbol{\beta}) - \tau \big] \big[\int_0^Y G(s, \mathbf{x}) ds \big]^{-1}$ for $\mathbf{z}^T = (1, \mathbf{x}^T)$, and $\mathbb{G} = \{ G(t, \mathbf{x}) : ||\mathbf{x}|| \le L, 0 \le t \le t_1; c_1 \le G(t, \mathbf{x}) \le 1,$ for all $\mathbf{x}$ and $0 \le t \le t_0; G(t, \mathbf{x})$ is nonincreasing in t, $\forall\ \mathbf{x};\ |G(t, \mathbf{x}_1) - G(t, \mathbf{x}_2)| \le c_2 ||\mathbf{x}_1 - \mathbf{x}_2||,\ \forall\ t,$ for a positive constant c_2. It can be checked that for the local Kaplan-Meier estimator defined in (5), $1 - \widehat{G}_2(s, \mathbf{x})$ is a distribution function. Thus $\widehat{G}_2 \in \mathbb{G}$. The lemma below establishes that the size of $\mathbb{F}$ is not too large.

Lemma 5.1 $\mathbb{F}$ *is P-Donsker.*

Proof. Given any $\epsilon > 0$, let $N(\epsilon, \mathbb{G}, L_2(Q))$ denote the ϵ-covering number for $\mathbb{G}$, where Q is an arbitrary probability measure. We cover the set $\{\mathbf{x} : ||\mathbf{x}|| \le L\}$ with N_1 balls with $L_2(Q)$ radius ϵ such that $N_1 = \frac{c_3}{\epsilon^d}$, for some constant $c_3 > 0$. Denote the center of these N_1 balls by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_1}$. Since $\mathbb{G}_i = \{G(t, \mathbf{x}_i) \in \mathbb{G}\}$ is a monotone class of functions, $\forall\ \mathbf{x}_i$, we have $N\{\epsilon, \mathbb{G}_i, L_2(Q)\} \le \exp(c_4\epsilon^{-1})$, where $c_4 > 0$ is a constant. Now, $\forall\ G(t, \mathbf{x}) \in \mathbb{G}$, we can find $1 \le i \le N_1$ such that $||\mathbf{x} - \mathbf{x}_i|| \le \epsilon$. Note that $G(t, \mathbf{x}_i)$ is within one of the balls that cover $\mathbb{G}_i$. Denote the center of this ball by G^*. Then $G(t, \mathbf{x})$ is covered by the ball with the center G^* and radius $(1+c_2)\epsilon$ by the Lipschitz condition. Thus $N\{(1+c_2)\epsilon, \mathbb{G}, L_2(Q)\} \le \frac{c_3}{\epsilon^d}\exp(c_4\epsilon^{-1})$. Hence $N(\epsilon, \mathbb{G}, L_2(Q)) \le \frac{K_1}{\epsilon^d}\exp(K_2\epsilon^{-1})$, for some positive constants K_1 and K_2.

Next, define $\mathbb{F}_1 = \left\{1/\int_0^y G(s, \mathbf{x})ds : \ G(s, \mathbf{x}) \in \mathbb{G}\right\}$. Note that $\mathbb{G}$ can be covered by N balls of $L_2(Q)$ radius ϵ, where $N = N(\epsilon, \mathbb{G}, L_2(Q))$. For any $G(s, x) \in \mathbb{G}$, it is covered by one of the N balls. Consider any $\widetilde{G}(s, x) \in \mathbb{G}$ such that $\widetilde{G}(s, x)$ is covered by the same ball. Then applying Hölder's inequality and Condition (C6),

$$
\begin{aligned}
& E\Big[\frac{1}{\int_0^Y G(s,\mathbf{x})ds} - \frac{1}{\int_0^Y \widetilde{G}(s,\mathbf{x})ds}\Big]^2 = E\Big[\frac{(\int_0^Y (G(s,\mathbf{x}) - \widetilde{G}(s,\mathbf{x}))ds)^2}{(\int_0^Y G(s,\mathbf{x})ds)^2(\int_0^Y \widetilde{G}(s,\mathbf{x})ds)^2}\Big] \\
\le\ & \frac{1}{c_1^4}E\Big[\frac{(\int_0^\infty I(s<Y)(G(s,\mathbf{x}) - \widetilde{G}(s,\mathbf{x}))ds)^2}{\min(Y,t_0)^4}\Big] \\
\le\ & \frac{1}{c_1^4}E\Big[\frac{(\int_0^\infty I(s<Y)ds)(\int_0^\infty (G(s,\mathbf{x}) - \widetilde{G}(s,\mathbf{x}))^2 ds}{\min(Y,t_0)^4}\Big] \\
\le\ & \frac{1}{c_1^4}E\Big[E(Y\min(Y,t_0)^{-4}|\mathbf{x})E\Big(\int_0^\infty (G(s,\mathbf{x}) - \widetilde{G}(s,\mathbf{x}))^2 ds\Big)\Big] \le \frac{c_5\epsilon^2}{c_1^4}
\end{aligned}
$$

Therefore, we have $N(\sqrt{c_5}c_1^{-2}\epsilon, \mathbb{F}_1, L_2(Q)) \le N$. This implies that $\mathbb{F}_1$ has bounded uniform entropy integral.

Finally, let $\mathbb{F}_2 = \left\{\delta\mathbf{z}[I(Y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] : \ \beta \in \mathbb{B}\right\}$. Then $\mathbb{F} = \mathbb{F}_1 \cdot \mathbb{F}_2$. Since $\mathbb{F}_2$ is a VC graph class of functions, it has bounded uniform entropy integral. Thus the product class $\mathbb{F}$ has bounded uniform entropy integral, which implies that $\mathbb{F}$ is P-Donsker. □

Lemma 5.2 *Under conditions (C1)-(C8), if* $h_n = n^{-\alpha}$ *with* $0 < \alpha < \min(1/\nu, 1/p)$, *then*

(i) $\sup_s \sup_{\mathbf{x}} |\hat{G}(s|\mathbf{x}) - G_0(s|\mathbf{x})| = O\left(\{\log n/(nh_n^p)\}^{1/2} + h_n^\nu\right)$ *a.s.;*

(ii) $\hat{G}(s|\mathbf{x}) - G_0(s|\mathbf{x}) = -\sum_{j=1}^n B_{nj}(\mathbf{x})\xi(Y_j^*, \delta_j, s, \mathbf{x}) + O\left(\{\log n/(nh_n^p)\}^{3/4} + h_n^\nu\right)$ *a.s.*

for $s < s_0$ *such that* $\inf_{\mathbf{x}}\{1 - F_0(s_0|\mathbf{x})\}G_0(s_0|\mathbf{x}) > 0$, *where*

$$\xi(V_j^*, \delta_j, s, \mathbf{x}) = G_0(s|\mathbf{x}) \left[\int_0^{\min(Y_j^*, s)} \frac{dG_0(u|\mathbf{x})}{\{1 - F_0(u|\mathbf{x})\}G_0^2(u|\mathbf{x})} + \frac{I(V_j^* \le s, \delta_j = 0)}{\{1 - F_0(V_j^*|\mathbf{x})\}G_0(Y_j^*|\mathbf{x})} \right],$$

and $V_j^* = \min(V_j, C_j)$.

Proof. The proof follows directly from the same arguments for proving Theorems 2.2 and 2.3 of Gonzalez-Manteiga & Cadarso-Suarez (1994). The main difference is that the bias influence and the variance influence are h^ν and $(nh_n^p)^{-1}$ for a νth order kernel function for p-dimensional covariates. $\square$

Proof of Theorem 2.1(2). We'll check the conditions of Theorem 1 of Chen et al. (2003). Their condition (1.3) is straightforward and condition (1.4) follows from Corollary 2.1 of Dabrowska (1989). By the subgradient condition $||\mathbf{M}_n(\widehat{\boldsymbol{\beta}}, \widehat{G}_2)|| = o_p(n^{-1/2})$, their condition (1.1) is satisfied. By Lemma 5.1, $\mathbb{F} = \left\{ m(\mathbf{w}, \boldsymbol{\beta}, G) : \ \boldsymbol{\beta} \in \mathbb{B}, G \in \mathbb{G} \right\}$ is P-Donsker, thus is also P-Glivenko-Cantelli (van der Vaart and Wellner, 1996). This implies their condition (1.5). It remains to check their condition (1.2). Let $F_{\widetilde{T}}(\cdot|\mathbf{x})$ be the conditional CDF of $\widetilde{T}$ given $\mathbf{x} = \mathbf{x}$; similarly let $F_\epsilon(\cdot|\mathbf{x})$ be the conditional CDF of ϵ. Then

$$\begin{aligned}
\inf_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \ge \delta} ||\mathbf{M}(\boldsymbol{\beta}, G_0)|| &= \inf_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \ge \delta} ||\mathbf{M}(\boldsymbol{\beta}, G_0) - \mathbf{M}(\boldsymbol{\beta}_0, G_0)|| \\
&= \inf_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \ge \delta} \Big|\Big| E\Big[\frac{\mathbf{z}}{\mu(\mathbf{x})} [F_{\widetilde{T}}(\mathbf{z}^T\boldsymbol{\beta}|\mathbf{x}) - F_{\widetilde{T}}(\mathbf{z}^T\boldsymbol{\beta}_0|\mathbf{x}) \Big] \Big|\Big| \\
&= \inf_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \ge \delta} \Big|\Big| E\Big[\frac{\mathbf{z}}{\mu(\mathbf{x})} [F_\epsilon(\mathbf{z}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\mathbf{x}) - F_\epsilon(0|\mathbf{x}) \Big] \Big|\Big| \\
&= \inf_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \ge \delta} \Big|\Big| E\Big[\frac{\mathbf{z}\mathbf{z}^T}{\mu(\mathbf{x})} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) f_\epsilon(\xi|\mathbf{x}) \Big] \Big|\Big| \\
&\ge C\delta\lambda_{\min}\big[E(\mathbf{z}\mathbf{z}^T/\mu(\mathbf{x}))\big] \ge C^*\delta,
\end{aligned}$$

where ξ is between 0 and $\mathbf{z}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, C and C^* are finite positive constants. The last equality follows from the mean value theorem, the second last inequality uses condition (C5) and the last inequality uses condition (C7). Thus their condition (1.2) is also satisfied. $\square$

Proof of Theorem 2.2(2). Recall that

$$\mathbf{M}(\boldsymbol{\beta}, G_0) = E\Big[\frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty [I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x}) \, \mathrm{d}y \Big] = E\Big[\frac{\mathbf{z}}{\mu(\mathbf{x})} [F_\epsilon\{\mathbf{z}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\mathbf{x}\} - \tau] \Big].$$

Therefore, we have $\mathbf{\Gamma}_1(\boldsymbol{\beta}_0, G_0) = \frac{\partial \mathbf{M}(\boldsymbol{\beta}, G_0)}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = E\Big[\frac{\mathbf{z}\mathbf{z}^T}{\mu(\mathbf{x})} f_\epsilon(0|\mathbf{x})\Big]$. And the pathwise derivative in the direction $G - G_0$ is

$$
\begin{aligned}
& \mathbf{\Gamma}_2(\boldsymbol{\beta}, G_0)[G - G_0] \\
= \;& \lim_{a \to 0} \frac{[\mathbf{M}(\boldsymbol{\beta}, G_0 + a(G - G_0)) - \mathbf{M}(\boldsymbol{\beta}, G_0)]}{a} \\
= \;& \lim_{a \to 0} \frac{1}{a} \mathrm{E}\Big\{ \frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty \Big[\frac{\int_0^y G_0(s|\mathbf{x})\,\mathrm{d}s}{\int_0^y (G_0(s|\mathbf{x}) + a(G(s|\mathbf{x}) - G_0(s|\mathbf{x})))\,\mathrm{d}s} - 1 \Big] [I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\} \\
= \;& -\mathrm{E}\Big\{ \frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty \frac{\int_0^y (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))\,\mathrm{d}s}{\int_0^y G_0(s|\mathbf{x})\,\mathrm{d}s} [I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\}.
\end{aligned}
$$

We check the conditions of Theorem 2 in Chen et al. (2003). Note that (2.1) has been proved earlier, (2.2) is evident from the above calculation, (2.4) is satisfied by the definition of $\mathbb{G}$, and (2.5) is implied by Lemma 5.1. We'll check conditions (2.3) and (2.6).

Condition (2.3). Consider any $\boldsymbol{\beta} \in \mathbb{B}$ and $G \in \mathbb{G}$ such that $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \le \delta_n$ and $||G - G_0|| \le \delta_n$, where $\delta_n = o(1)$ and for a function $G(t, x)$, $||G(t, \mathbf{x})|| = \sup_t \sup_{\mathbf{x}} |G(t, \mathbf{x})|$ denotes the sup-norm. We have

$$
\begin{aligned}
& ||\mathbf{M}(\boldsymbol{\beta}, G) - \mathbf{M}(\boldsymbol{\beta}, G_0) - \mathbf{\Gamma}_2(\boldsymbol{\beta}, G_0)[G - G_0]|| \\
= \;& \Big|\Big| \mathrm{E}\Big\{ \frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty \frac{\int_0^y (G_0(s|\mathbf{x}) - G(s|\mathbf{x}))\,\mathrm{d}s}{\int_0^y G(s|\mathbf{x})\,\mathrm{d}s} [I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \\
& + \frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty \frac{\int_0^y (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))\,\mathrm{d}s}{\int_0^y G_0(s|\mathbf{x})\,\mathrm{d}s} [I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\} \Big|\Big| \\
= \;& \Big|\Big| \mathrm{E}\Big\{ \frac{\mathbf{z}}{\mu(\mathbf{x})} \int_0^\infty \frac{(\int_0^y (G_0(s|\mathbf{x}) - G(s|\mathbf{x}))\,\mathrm{d}s)^2}{\int_0^y G(s|\mathbf{x})\,\mathrm{d}s \int_0^y G_0(s|\mathbf{x})\,\mathrm{d}s} [I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\} \Big|\Big| \\
\le \;& C \Big|\Big| \mathrm{E}\Big\{ \int_0^\infty \frac{(\int_0^\infty I(s \le y)(G_0(s|\mathbf{x}) - G(s|\mathbf{x}))\,\mathrm{d}s)^2}{\int_0^y G(s|\mathbf{x})\,\mathrm{d}s \int_0^y G_0(s|\mathbf{x})\,\mathrm{d}s} f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\} \Big|\Big| \\
\le \;& \frac{C}{c_1^2} \Big|\Big| \mathrm{E}\Big\{ \int_0^\infty \frac{(\int_0^\infty I(s \le y)(G_0(s|\mathbf{x}) - G(s|\mathbf{x}))\,\mathrm{d}s)^2}{min(y, t_0)^2} f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\} \Big|\Big| \\
\le \;& \frac{C}{c_1^2} \Big|\Big| \mathrm{E}\Big\{ \int_0^\infty \frac{y^2 (\int_0^\infty (G_0(s|\mathbf{x}) - G(s|\mathbf{x}))\,\mathrm{d}s)^2}{min(y, t_0)^2} f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y \Big\} \Big|\Big| \\
\le \;& \frac{C}{c_1^2} \Big|\Big| \mathrm{E}\Big[\int_0^\infty (G_0(s|\mathbf{x}) - G(s|\mathbf{x}))^2\,\mathrm{d}s \Big] \Big|\Big| \le \frac{Ct_1}{c_1^2} ||G - G_0||^2,
\end{aligned}
$$

where C stands for a finite positive constant. Thus (2.3)(i) is satisfied. Furthermore,

$$
\begin{aligned}
& ||\mathbf{\Gamma}_2(\boldsymbol{\beta}, G_0)[G-G_0] - \mathbf{\Gamma}_2(\boldsymbol{\beta}_0, G_0)[G-G_0]|| \\
= \;& \Big|\Big|\mathrm{E}\Big\{\frac{\mathbf{z}}{\mu(\mathbf{x})}\int_0^\infty \frac{\int_0^y (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))\,\mathrm{d}s}{\int_0^y G_0(s|\mathbf{x})\,\mathrm{d}s}[I(y \le \mathbf{z}^T\boldsymbol{\beta}) - I(y \le \mathbf{z}^T\boldsymbol{\beta}_0)] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y\Big\}\Big|\Big| \\
\le \;& C\Big|\Big|\mathrm{E}\Big\{\int_0^\infty \frac{y(\int_0^\infty (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))^2\,\mathrm{d}s)^{1/2}}{c_1 \min(y, t_0)}|I(y \le \mathbf{z}^T\boldsymbol{\beta}) - I(y \le \mathbf{z}^T\boldsymbol{\beta}_0)| f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y\Big\}\Big|\Big| \\
\le \;& \frac{C}{c_1}\Big|\Big|\mathrm{E}\Big\{\big(\int_0^\infty (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))^2\,\mathrm{d}s\big)^{1/2}\int_0^\infty |I(y \le \mathbf{z}^T\boldsymbol{\beta}) - I(y \le \mathbf{z}^T\boldsymbol{\beta}_0)| f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y\Big\}\Big|\Big| \\
\le \;& C^*||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \cdot \Big|\Big|\mathrm{E}\Big[\big(\int_0^\infty (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))^2\,\mathrm{d}s\big)^{1/2}\Big]\Big|\Big| \\
\le \;& C^* t_1^{1/2}\delta_n\Big|\Big|\Big[\mathrm{E}\big(\int_0^\infty (G(s|\mathbf{x}) - G_0(s|\mathbf{x}))^2\,\mathrm{d}s\big)\Big]^{1/2}\Big|\Big| \le C^* t_1^{1/2}\delta_n ||G - G_0|| = o_p(1)\delta_n.
\end{aligned}
$$

Thus (2.3)(ii) is satisfied.

Condition (2.6). We consider the almost sure i.i.d. representation in Lemma 5.2:

$$
\begin{aligned}
\widehat{G}_2(s|\mathbf{x}) - G_0(s|\mathbf{x}) &= -\sum_{j=1}^n B_{nj}(\mathbf{x})\xi(V_j^*, \delta_j, s, \mathbf{x}) + O_p\big((\log n/(nh_n^p))^{3/4} + h_n^\nu\big) \\
&= -n^{-1}h_n^{-p}\sum_{j=1}^n \frac{K\left(\frac{\mathbf{x}-\mathbf{x}_j}{h_n}\right)}{n^{-1}h_n^{-p}\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_n}\right)}\xi(V_j^*, \delta_j, s, \mathbf{x}) + O_p\big((\log n/(nh_n^p))^{3/4} + h_n^\nu\big),
\end{aligned}
$$

where $V_j^* = min(C_j, V_j)$ and

$$
\xi(V_j^*, \delta_j, s, \mathbf{x}) = G_0(s|\mathbf{x})\left[\int_0^{\min(V_j^*, s)} \frac{dG_0(u|\mathbf{x})}{(1 - F_0(u|\mathbf{x}))G_0^2(u|\mathbf{x})} + \frac{I(V_j^* \le s, \delta_j = 0)}{(1 - F_0(V_j^*|\mathbf{x}))G_0(V_j^*|\mathbf{x})}\right]
$$

are independent random variables with mean zero and finite variance for any given s and $\mathbf{x}$. Now by standard change of variables and Taylor expansion,

$$
\mathbf{\Gamma}_2(\boldsymbol{\beta}_0, G_0)[\widehat{G}_2 - G_0] = n^{-1}\sum_{j=1}^n \phi_2(V_j^*, \delta_j, \mathbf{x}_j) + o_p(n^{-1/2}),
$$

where

$$
\phi_2(V_j^*, \delta_j, \mathbf{x}_j) = \frac{\mathbf{z}_j}{\mu(\mathbf{x}_j)}\int_0^\infty \frac{\int_0^y \xi(V_j^*, \delta_j, s, \mathbf{x}_j)\,\mathrm{d}s}{\int_0^y G_0(s|\mathbf{x}_j)\,\mathrm{d}s}[I(y \le \mathbf{z}_j^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x}_j)\,\mathrm{d}y.
$$

Since $E(\xi(V_j^*, \delta_j, s, \mathbf{x}_j)|\mathbf{x}_j = \mathbf{x}) = 0$ (Gonzalez-Manteiga and Cadarso-Suarez, 1994), $\phi(V_j^*, \delta_j, \mathbf{x}_j)$

are i.i.d. with mean zero.

Recall that $\mathbf{M}_n(\boldsymbol{\beta}_0, G_0) = n^{-1}\sum_{i=1}^n m(\mathbf{W}_i, \boldsymbol{\beta}_0, G_0)$ with $m(\mathbf{W}_i, \boldsymbol{\beta}_0, g_0) = \delta_i \mathbf{z}_i \frac{[I(Y_i \le \mathbf{z}_i^T \boldsymbol{\beta}) - \tau]}{\int_0^{Y_i} G_0(s|\mathbf{x}_i)\,\mathrm{d}s}$ being independent mean 0 random vectors. Therefore,

$$\sqrt{n}\{\mathbf{M}_n(\boldsymbol{\beta}_0, G_0) + \boldsymbol{\Gamma}_2(\boldsymbol{\beta}_0, G_0)[\widehat{G}_2 - G_0]\} \to N(\mathbf{0}, \mathbf{V}_2),$$

where $\mathbf{V}_2 = \mathrm{Cov}(m(\mathbf{W}_i, \boldsymbol{\beta}_0, G_0) + \phi_2(V_j^*, \delta_j, \mathbf{x}_j))$. In conclusion, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0)^{-1}\mathbf{V}_2\boldsymbol{\Gamma}_1(\boldsymbol{\beta}_0, G_0)^{-1}).$$

Remark. In the case where the censoring variable C does not depend on the covariates (Theorem 2.2(1)), the above argument proceeds similarly. Using the almost sure i.i.d. representation of the Kaplan-Meier estimator (Theorem 1, Lo and Singh, 1986), we have

$$\widehat{G}_1(s) - G_0(s) \quad = \quad -n^{-1}\sum_{j=1}^n \xi_1(V_j^*, \delta_j, s) + O_p\big(n^{-3/4}(\log n)^{3/4}\big),$$

where $V_j^* = \min(C_j, V_j)$, and

$$\xi_1(V_j^*, \delta_j, s) = G_0(s)\left[\int_0^{\min(V_j^*, s)} \frac{dG_0(u)}{\{1 - F_0(u)\}G_0^2(u)} + \frac{I(V_j^* \le s, \delta_j = 0)}{\{1 - F_0(V_j^*)\}G_0(V_j^*)}\right]$$

are independent with mean zero and finite variance for any given s. And we have

$$\boldsymbol{\Gamma}_2(\boldsymbol{\beta}_0, G_0)[\widehat{G}_1 - G_0] = n^{-1}\sum_{j=1}^n \phi_1(V_j^*, \delta_j) + o_p(n^{-1/2}),$$

where $\phi_1(V_j^*, \delta_j) = \int \frac{\mathbf{x}h(\mathbf{x})}{\mu(\mathbf{x})} \int_0^\infty \frac{\int_0^y \xi_1(V_j^*, \delta_j, s)\,\mathrm{d}s}{\int_0^y G_0(s)\,\mathrm{d}s}[I(y \le \mathbf{z}^T\boldsymbol{\beta}) - \tau] f_{\widetilde{T}}(y|\mathbf{x})\,\mathrm{d}y\mathrm{d}\mathbf{x}$, and the outer-layer integral denotes a p-dimensional integral. □

Proof of Theorem 2.3. Let $T_{n2}^* = \frac{1}{n(n-1)\lambda_n^p}\sum_{i=1}^n\sum_{j=1, j\ne i}^n K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\lambda_n}\right) e_i e_j$, where $e_i = \delta_i[I(Y_i \le \mathbf{z}_i^T\boldsymbol{\beta}_0) - \tau] w_0^{-1}(Y_i|\mathbf{x}_i)$. We will show that $n\lambda_n^{p/2}(T_{n2} - T_{n2}^*) \to 0$ in probability. Then the theorem holds because $n\lambda_n^{p/2} T_{n2}^*$ has the desirable asymptotic distribution following the central limit theorem of Hall (1984) for a second-order degenerate U-statistic.

Note that we can write $\widehat{e}_i = e_i + \xi_i + \eta_i$, where $\xi_i = \delta_i[I(Y_i \le \mathbf{z}_i^T\widehat{\boldsymbol{\beta}}) - I(Y_i \le \mathbf{z}_i^T\boldsymbol{\beta}_0)]\widehat{w}^{-1}(Y_i|\mathbf{x}_i)$ and $\eta_i = \delta_i[I(Y_i \le \mathbf{z}_i^T\boldsymbol{\beta}_0) - \tau][\widehat{w}^{-1}(Y_i|\mathbf{x}_i) - w_0^{-1}(Y_i|\mathbf{x}_i)]$, where $\widehat{w}(Y_i|\mathbf{x}_i) = \int_0^{Y_i} \widehat{G}_2(s|\mathbf{x}_i)$ and

$w_0(Y_i|\mathbf{x}_i) = \int_0^{Y_i} G_0(s|\mathbf{x}_i)$. Thus

$$\begin{aligned} T_{n2} &= T_{n2}^* + \frac{1}{n(n-1)\lambda_n^p} \sum_{j \neq i} e_i \xi_j + \frac{1}{n(n-1)\lambda_n^p} \sum_{j \neq i} e_i \eta_j + \frac{1}{n(n-1)\lambda_n^p} \sum_{j \neq i} \xi_i \xi_j \\ & \quad + \frac{1}{n(n-1)\lambda_n^p} \sum_{j \neq i} \eta_i \eta_j + \frac{1}{n(n-1)\lambda_n^p} \sum_{j \neq i} \xi_i \eta_j = T_{n2}^* + \sum_{t=1}^{5} D_t, \end{aligned}$$

where the definition of D_t, $t = 1, \ldots, 5$, should be clear from the context. To prove the theorem, it suffices to establish that $n\lambda_n^{p/2} D_t = o_p(1)$, $t = 1, \ldots, 5$.

Let $Q(\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{\mu(\mathbf{x})} P_{\widetilde{T}|\mathbf{x}}(\widetilde{T} \leq \mathbf{z}^T \boldsymbol{\beta})$. From the calculations in Section 3.1, $Q(\mathbf{x}, \boldsymbol{\beta}) = E\Big[\delta_i I(Y_i \leq \mathbf{z}_i^T \boldsymbol{\beta}) w_0^{-1}(Y_i|\mathbf{x}_i) \Big| \mathbf{x}_i = \mathbf{x}\Big]$. We can write

$$\begin{aligned} & n\lambda_n^{p/2} D_1 \\ = \; & [(n-1)\lambda_n^{p/2}]^{-1} \sum_{j \neq i} K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\lambda_n}\right) e_i \Big[\delta_j I(Y_j \leq \mathbf{z}_j^T \widehat{\boldsymbol{\beta}}) w_0^{-1}(Y_j|\mathbf{x}_j) - \delta_j I(Y_j \leq \mathbf{z}_j^T \boldsymbol{\beta}_0) w_0^{-1}(Y_j|\mathbf{x}_j) \\ & - Q(\mathbf{x}_j, \widehat{\boldsymbol{\beta}}) + Q(\mathbf{x}_j, \boldsymbol{\beta}_0)\Big] + [(n-1)\lambda_n^{p/2}]^{-1} \sum_{j \neq i} K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\lambda_n}\right) e_i \Big[Q(\mathbf{x}_j, \widehat{\boldsymbol{\beta}}) - Q(\mathbf{x}_j, \boldsymbol{\beta}_0)\Big] \\ & + [(n-1)\lambda_n^{p/2}]^{-1} \sum_{j \neq i} K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\lambda_n}\right) e_i \delta_j \Big[I(Y_j \leq \mathbf{z}_j^T \widehat{\boldsymbol{\beta}}) - I(Y_j \leq \mathbf{z}_j^T \boldsymbol{\beta}_0)\Big] \\ & \times \big[\widehat{w}^{-1}(Y_j|\mathbf{x}_j) - w_0^{-1}(Y_j|\mathbf{x}_j)\big] = D_{11} + D_{12} + D_{13}, \end{aligned}$$

where the definition of D_{1i}, $i = 1, 2, 3$, is clear from the context. To prove that $D_{11} = o_p(1)$, it suffices to show that $\sup_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq cn^{-1/2}} |D_{11}(\boldsymbol{\beta}, \boldsymbol{\beta}_0)| = o_p(1)$ for any positive constant c, where $D_{11}(\boldsymbol{\beta}, \boldsymbol{\beta}_0) = [(n-1)\lambda_n^{p/2}]^{-1} \sum_{j \neq i} H_n(\mathbf{W}_i, \mathbf{W}_j, \boldsymbol{\beta}, \boldsymbol{\beta}_0)$ and $H_n(\mathbf{W}_i, \mathbf{W}_j, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = K\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{\lambda_n}\right) e_i \Big[\delta_j I(Y_j \leq \mathbf{z}_j^T \boldsymbol{\beta}) w_0^{-1}(Y_j|\mathbf{x}_j) - \delta_j I(Y_j \leq \mathbf{z}_j^T \boldsymbol{\beta}_0) w_0^{-1}(Y_j|\mathbf{x}_j) - Q(\mathbf{x}_j, \boldsymbol{\beta}) + Q(\mathbf{x}_j, \boldsymbol{\beta}_0)\Big]$. We observe that $n^{-1}\lambda_n^{p/2} D_{11}(\boldsymbol{\beta}, \boldsymbol{\beta}_0) = [n(n-1)]^{-1} \sum_{j \neq i} H_n(\mathbf{W}_i, \mathbf{W}_j, \boldsymbol{\beta}, \boldsymbol{\beta}_0)$ is a second-order degenerate U-statistic and that the class of functions $\{H_n(W_i, W_j, \boldsymbol{\beta}, \boldsymbol{\beta}_0) : \boldsymbol{\beta} \in \mathbb{B}\}$ is Euclidean with a constant envelope function. Applying the result on uniform convergence of U-statistics indexed by parameters (Lemma 1, Zheng, 1998), we have

$$\sup_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq cn^{-1/2}} |n^{-1}\lambda_n^{p/2} D_{11}(\boldsymbol{\beta}, \boldsymbol{\beta}_0)| = O_p(n^{-1-\gamma/4}), \quad \forall\; 0 < \gamma < 1.$$

So $\sup_{||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq cn^{-1/2}} |D_{11}(\boldsymbol{\beta}, \boldsymbol{\beta}_0)| = O(n\lambda_n^{-p/2}) O_p(n^{-1-\gamma/4}) = O_p(\lambda_n^{-p/2} n^{-\gamma/4}) = o_p(1)$. Next, let $D_{12}(\boldsymbol{\beta},\boldsymbol{\beta}_0) = [(n-1)\lambda_n^{p/2}]^{-1}\sum_{j\neq i} K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right) e_i\Big[Q(\mathbf{x}_j,\boldsymbol{\beta}) - Q(\mathbf{x}_j,\boldsymbol{\beta}_0)\Big]$. Let $Q_1(\boldsymbol{\beta}) = \frac{1}{\mu(\mathbf{x})}\frac{d}{d\xi}P_{\widetilde{T}|\mathbf{x}}(\widetilde{T}\leq\xi)\Big|_{\xi=\mathbf{z}^T\boldsymbol{\beta}}$ and $Q_2(\boldsymbol{\beta}) = \frac{1}{\mu(\mathbf{x})}\frac{d^2}{d\xi^2}P_{\widetilde{T}|\mathbf{x}}(\widetilde{T}\leq\xi)\Big|_{\xi=\mathbf{z}^T\boldsymbol{\beta}}$. By Taylor expansion,

$$\begin{aligned} D_{12}(\boldsymbol{\beta},\boldsymbol{\beta}_0) &= [(n-1)\lambda_n^{p/2}]^{-1}\sum_{j\neq i} K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right) e_i Q_1(\boldsymbol{\beta}_0)(\mathbf{z}_j^T\boldsymbol{\beta}-\mathbf{z}_j^T\boldsymbol{\beta}_0) \\ &\quad + [(n-1)\lambda_n^{p/2}]^{-1}\sum_{j\neq i} K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right) e_i Q_2(\boldsymbol{\beta}^*)(\mathbf{z}_j^T\boldsymbol{\beta}-\mathbf{z}_j^T\boldsymbol{\beta}_0)^2, \end{aligned}$$

where $\boldsymbol{\beta}^*$ is between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$. Following similar lines as in Wang (2008), it can be shown that $\sup_{||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\leq cn^{-1/2}}|D_{12}(\boldsymbol{\beta},\boldsymbol{\beta}_0)| = O_p(\lambda_n^{p/2}) = o_p(1)$. Thus $D_{12} = o_p(1)$. To evaluate D_{13}, we note that we can assume $Y_j \leq \max\{\mathbf{z}_j^T\widehat{\boldsymbol{\beta}}, \mathbf{z}_j^T\boldsymbol{\beta}_0\}$ otherwise $I(Y_j\leq\mathbf{z}_j^T\widehat{\boldsymbol{\beta}}) - I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}_0)$ is zero, then by Lemma 5.2,

$$\begin{aligned} |D_{13}| &\leq C[(n-1)\lambda_n^{p/2}]^{-1}\sum_{j=1}^n\Big|\sum_{i=1,i\neq j}^n K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right)e_i\Big|\Big|I(Y_j\leq\mathbf{z}_j^T\widehat{\boldsymbol{\beta}}) - I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}_0)\Big| \\ &\quad \widehat{w}^{-1}(Y_j|\mathbf{x}_j)w_0^{-1}(Y_j|\mathbf{x}_j)|Y_j|\sup_s\sup_{\mathbf{x}}\left|\widehat{G}(s|\mathbf{x}) - G_0(s|\mathbf{x})\right| \\ &= C[(n-1)\lambda_n^{p/2}]^{-1}O_p\left((\log n)^{1/2}n^{-1/2}h_n^{-p/4}+h_n^{\nu}\right) \\ &\quad \sum_{j=1}^n\Big|\sum_{i=1,i\neq j}^n K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right)e_i\Big|\Big|I(Y_j\leq\mathbf{z}_j^T\widehat{\boldsymbol{\beta}}) - I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}_0)\Big| \end{aligned}$$

Let $G_{13}(\boldsymbol{\beta},\boldsymbol{\beta}_0) = \sum_{j=1}^n\Big|\sum_{i=1,i\neq j}^n K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right)e_i\Big|\Big|I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}) - I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}_0)\Big|$. Then by Hölder's inequality,

$$\begin{aligned} & E\Big[\sup_{||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\leq cn^{-1/2}}G_{13}(\boldsymbol{\beta},\boldsymbol{\beta}_0)\Big] \\ \leq\ & \sum_{j=1}^n\Big\{E\big[\sum_{i=1,i\neq j}^n K\left(\frac{\mathbf{x}_i-\mathbf{x}_j}{\lambda_n}\right)e_i\big]^2\Big\}^{1/2}\Big\{E\big[\sup_{||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\leq cn^{-1/2}}\big|I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}) - I(Y_j\leq\mathbf{z}_j^T\boldsymbol{\beta}_0)\big|\big]\Big\}^{1/2} \\ =\ & O(n)O(n^{1/2}\lambda_n^{p/2})O(n^{-1/4}) = O(n^{5/4}\lambda_n^{p/2}). \end{aligned}$$

Thus $G_{13}(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta}_0) = O_p(n^{5/4}\lambda_n^{p/2})$. And this implies

$$|D_{13}| \leq O_p(n^{-1/4}h_n^{-p/4}(\log n)^{1/2}) + O(n^{1/4}h_n^{\nu}) = o_p(1),$$

from the conditions on h_n in Theorem 2.2. Putting the above together, we have $D_1 = o_p(1)$. Similarly, we can prove that $D_t = o_p(1)$, for $t = 2,3,4,5$. □

References

Asgharian, M., M'Lan, C. E. & Wolfson, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association* **97**, 201–209.

Asgharian, M. & Wolfson, D. B. (2005). Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics* **32**, 2109–2131.

Bang, H. & Tsiatis, A. (2002). Median regression with censored cost data. *Biometrics* **58**, 643–649.

Chen, X., Linton, O. & Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71**, 1591–1608.

Dabrowska, D. M. (1989). Uniform consistency of kernel conditional kaplan-meier estimate. *The Annals of Statistics* **17**, 1157–1167.

de Uña Álvarez, J. (2002). Product-limit estimation for length-biased censored data. *Test* **11**, 109–125.

de Uña Álvarez, J. & Iglesias-Pérez, M. C. (2010). Nonparametric estimation of a conditional distribution from length-biased data. *Annals of the Institute of Statistical Mathematics* **62**, 323–341.

de Uña Álvarez, J., Otero-Giráldez, M. S. & Álvarez-Llorente, G. (2003). Estimation under length-bias and right-censoring: an application to unemployment duration analysis for married women. *Journal of Applied Statistics* **30**, 283–291.

Gonzalez-Manteiga, W. & Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized kaplan-meier estimator with some applications. *Journal of Nonparametric Statistics* **4**, 65–78.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* **14**, 1–16.

Honoré, B., Khan, S. & Powell, J. L. (2002). Quantile regression under random censoring. *Journal of Econometrics* **109**, 67–105.

Huang, C. Y. & Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika* **98**, 177–186.

Jiang, H., Fine, J. P. & Chappell, R. (2005). Semiparametric analysis of survival data with left truncation and dependent right censoring. *Biometrics* **61**, 567–575.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

Koenker, R. & Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association* **96**, 458–468.

Lo, S. H. & Singh, K. (1985). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probability Theory and Related Fields* **71**, 455–465.

Luo, X. & Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika* **96**, 873–886.

McKeague, I. W., Subramanian, S. & Sun, Y. (2001). Median regression and the missing information principle. *Journal of Nonparametric Statistics* **13**, 709–727.

Ning, J., Qin, J. & Shen, Y. (2010). Non-parametric tests for right-censored data with biased sampling. *Journal of the Royal Statistical Society Series B* **72**, 609–630.

Ning, J., Qin, J. & Shen, Y. (2011). Buckley-james-type estimator with right-censored and length-biased data. *Biometrics* , in press.

Peng, L. & Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103**, 637–649.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98**, 1001–1012.

Qin, J. & Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under cox model. *Biometrics* **66**, 382–392.

Shen, P. S. (2009). Hazards regression for length-biased and right-censored data. *Statistics and Probability Letters* **79**, 457–465.

Shen, Y., Ning, J. & Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association* **104**, 1192–1202.

Shin, Y. (2009). Length-bias correction in transformation models with supplementary data. *Econometric Reviews* **28**, 658–681.

Wang, H. & Wang, L. (2009). Locally weighted censored quantile regression. *Journal of American Statistical Association* **104**, 1117–1128.

Wang, L. (2008). Nonparametric test for checking lack-of-fit of quantile regression model under random censoring. *Canadian Journal of Statistics* **36**, 321–336.

Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* **86**, 130–143.

Wang, M. C. (1996). Hazards regression analysis for length-biased data. *Biometrika* **83**, 343–354.

Ying, Z., Jung, S. H. & Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90**, 178–184.

Zelen, M. (2005). Forward and backward reference times and length biased sampling: age specific models. *Lifetime Data Analysis* **10**, 325–334.

Zheng, J. X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* **14**, 123–138.

Table 1: Simulation results for Case 1 with homoscedastic normal error and covariate-independent censoring. The MSE is the estimated mean squared error and CovP is the empirical coverage probability of a 95% bootstrap confidence interval.

				100×Bias			100×MSE			100×CovP		
n	CP	τ	Method	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$
200	20	0.25	Naive	24.7	2.1	2.2	0.9	1.2	3.3	33.4	96.6	95.0
			IPW	-0.3	0.2	-0.3	2.0	2.6	9.0	93.4	95.2	94.6
			LIPW	-0.4	0.4	-0.7	2.1	2.6	9.6	92.8	95.2	94.8
		0.5	Naive	24.0	2.9	4.6	0.8	1.1	3.0	31.4	94.4	95.4
			IPW	-0.3	0.4	0.5	1.1	1.5	4.5	94.6	95.0	95.6
			LIPW	-0.6	0.8	1.1	1.1	1.4	4.8	94.8	96.2	96.4
	40	0.25	Naive	23.6	5.2	7.9	1.0	1.6	5.1	41.0	94.8	93.6
			IPW	-0.4	0.5	-0.8	2.2	3.0	9.7	93.2	94.6	93.8
			LIPW	-0.9	1.1	-0.6	2.2	3.0	10.0	94.2	96.4	95.2
		0.5	Naive	23.0	7.2	12.9	1.1	1.8	6.1	43.8	92.4	93.0
			IPW	-0.4	1.0	0.7	1.2	1.8	5.1	94.2	95.8	96.2
			LIPW	-1.0	1.6	0.7	1.2	1.8	5.1	95.8	96.6	96.8
400	20	0.25	Naive	24.4	2.1	2.6	0.5	0.7	1.7	8.6	95.4	95.8
			IPW	0.1	-0.4	-0.8	0.9	1.2	3.8	93.0	94.8	95.2
			LIPW	0.1	-0.4	-0.5	0.9	1.2	3.8	93.2	95.0	95.4
		0.5	Naive	23.9	3.4	3.8	0.4	0.5	1.7	5.6	93.6	94.2
			IPW	-0.4	0.4	0.4	0.5	0.8	2.3	95.0	95.6	95.8
			LIPW	-0.5	0.5	0.7	0.5	0.8	2.3	95.4	95.8	95.8
	40	0.25	Naive	23.2	5.5	9.5	0.5	0.8	2.6	15.8	90.6	91.0
			IPW	-0.2	-0.1	-1.0	1.0	1.4	4.3	92.4	94.6	94.0
			LIPW	-0.1	-0.2	-0.9	1.0	1.3	4.2	93.2	95.2	94.8
		0.5	Naive	22.5	8.5	14.3	0.5	0.8	3.5	13.2	86.2	88.4
			IPW	-0.5	0.6	0.7	0.6	0.9	2.7	94.6	94.8	94.8
			LIPW	-0.8	0.8	0.9	0.6	0.9	2.7	95.0	96.0	95.2

Table 2: Simulation results for Case 2 with heteroscedastic uniform error and covariate-independent censoring. The MSE is the estimated mean squared error and CovP is the empirical coverage probability of a 95% bootstrap confidence interval.

				100×Bias			100×MSE			100×CovP		
n	CP	τ	Method	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$
200	20	0.25	Naive	10.3	38.8	3.1	0.6	1.4	4.8	74.0	13.6	95.4
			IPW	0.7	0.6	-2.1	0.6	1.9	4.6	93.2	94.8	93.0
			LIPW	0.7	0.7	-1.5	0.6	1.9	4.6	94.0	95.2	93.8
		0.5	Naive	11.1	35.2	3.9	0.5	0.9	3.4	65.6	6.4	93.8
			IPW	-0.5	1.3	-0.8	0.6	1.7	4.4	94.2	94.2	96.0
			LIPW	-0.6	1.6	-0.0	0.6	1.7	4.5	94.6	94.6	95.2
	40	0.25	Naive	9.1	42.7	14.1	0.6	1.8	7.1	81.0	13.2	92.6
			IPW	0.9	0.6	-1.9	0.7	2.1	5.0	92.8	93.8	93.8
			LIPW	0.9	0.5	-2.2	0.7	2.2	5.0	93.8	94.8	95.4
		0.5	Naive	9.7	40.4	16.3	0.6	1.2	5.6	75.0	5.8	90.0
			IPW	-0.5	1.3	-0.8	0.7	2.2	4.9	93.4	95.6	95.0
			LIPW	-0.7	1.3	-0.1	0.7	2.1	5.0	93.8	94.8	96.0
400	20	0.25	Naive	10.2	38.3	5.6	0.3	0.7	2.5	54.6	1.0	93.4
			IPW	0.4	0.2	-0.9	0.3	0.9	2.3	93.6	96.2	94.8
			LIPW	0.3	0.3	-0.9	0.3	0.9	2.2	94.0	97.0	95.2
		0.5	Naive	11.4	34.8	5.1	0.2	0.4	1.6	40.4	0.6	93.2
			IPW	-0.3	0.4	0.1	0.3	0.8	2.3	94.4	96.0	97.0
			LIPW	-0.4	0.5	0.4	0.3	0.8	2.3	94.2	96.0	96.6
	40	0.25	Naive	8.8	43.0	17.4	0.3	0.8	3.4	68.0	1.2	84.6
			IPW	0.3	0.3	-0.4	0.3	1.0	2.5	93.4	95.6	94.8
			LIPW	0.2	0.5	-0.7	0.3	1.0	2.5	93.4	95.4	94.8
		0.5	Naive	10.0	40.3	17.5	0.3	0.6	2.4	57.2	0.4	82.4
			IPW	-0.3	0.5	0.3	0.3	1.0	2.6	93.8	94.4	96.4
			LIPW	-0.5	0.5	0.3	0.3	1.0	2.5	94.4	94.6	96.4

Table 3: Simulation results for Case 3 with heteroscedastic normal error and covariate-dependent censoring. The MSE is the estimated mean squared error and CovP is the empirical coverage probability of a 95% bootstrap confidence interval.

				100×Bias			100×MSE			100×CovP		
n	CP	τ	Method	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$
200	20	0.25	Naive	25.6	78.0	-20.9	1.4	2.6	10.9	44.0	0.8	90.8
			IPW	-1.1	0.4	-12.2	2.7	11.4	26.6	93.6	93.8	93.6
			LIPW	-1.0	-0.0	-6.3	2.6	11.3	26.4	93.6	94.4	94.4
		0.5	Naive	27.2	79.9	-29.0	1.2	2.5	12.0	36.0	0.4	86.2
			IPW	-0.7	1.2	-6.7	1.5	4.9	13.1	95.6	93.4	94.2
			LIPW	-0.6	0.2	-0.9	1.5	4.9	12.6	95.4	93.6	95.8
	40	0.25	Naive	24.4	83.3	-52.7	1.8	4.0	17.1	56.4	2.6	75.4
			IPW	-1.9	1.4	-25.2	2.7	12.4	29.3	94.2	92.2	**91.6**
			LIPW	-1.7	-0.2	-7.7	2.7	11.8	27.9	93.6	93.6	94.8
		0.5	Naive	26.9	89.7	-60.3	1.5	4.7	26.9	50.2	2.2	75.8
			IPW	-1.1	3.0	-21.6	1.7	5.3	14.3	94.8	94.6	**92.4**
			LIPW	-1.0	0.1	-2.1	1.6	5.3	13.5	96.0	95.2	95.6
400	20	0.25	Naive	25.9	78.5	-21.4	0.7	1.3	5.1	17.2	0.0	86.2
			IPW	-1.1	1.6	-7.6	1.3	4.2	12.1	94.4	95.2	94.2
			LIPW	-1.0	0.7	-2.2	1.2	4.2	12.4	94.0	95.4	94.2
		0.5	Naive	26.4	81.7	-28.2	0.6	1.2	5.8	8.6	0.0	77.6
			IPW	-0.2	1.2	-8.4	0.7	2.1	5.5	95.8	96.2	95.0
			LIPW	-0.2	0.2	-1.8	0.7	2.0	5.6	96.0	96.2	96.0
	40	0.25	Naive	24.9	84.2	-52.1	0.8	2.0	8.8	23.4	0.0	56.0
			IPW	-1.5	3.3	-19.9	1.3	4.7	12.9	93.8	95.2	**91.6**
			LIPW	-1.3	0.6	-2.9	1.3	4.6	12.9	94.2	96.0	95.2
		0.5	Naive	26.5	92.3	-56.9	0.8	2.6	17.5	16.2	0.0	58.8
			IPW	-0.4	3.2	-22.5	0.8	2.4	6.3	95.0	96.4	**86.0**
			LIPW	-0.2	-0.4	-2.2	0.8	2.3	6.4	96.0	96.4	95.2

Table 4: Estimates of Naive and IPW for the CSHA data set at $\tau = 0.1, \cdots, 0.9$

	IPW			Naive		
τ	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$	$\beta_{(0)}(\tau)$	$\beta_{(1)}(\tau)$	$\beta_{(2)}(\tau)$
0.1	6.10 (0.13)	-0.00 (0.57)	0.31 (0.34)	6.64 (0.10)	0.23 (0.12)	0.23 (0.15)
0.2	6.47 (0.15)	0.18 (0.23)	0.23 (0.20)	7.04 (0.10)	0.16 (0.13)	0.22 (0.13)
0.3	6.65 (0.13)	0.25 (0.16)	0.26 (0.18)	7.34 (0.09)	0.09 (0.11)	0.17 (0.13)
0.4	6.99 (0.16)	0.13 (0.18)	0.16 (0.19)	7.53 (0.07)	0.07 (0.09)	0.17 (0.10)
0.5	7.11 (0.12)	0.25 (0.14)	0.27 (0.16)	7.70 (0.09)	0.09 (0.11)	0.19 (0.11)
0.6	7.38 (0.10)	0.12 (0.12)	0.22 (0.13)	7.90 (0.11)	0.05 (0.12)	0.16 (0.14)
0.7	7.55 (0.08)	0.17 (0.10)	**0.27** (0.11)	8.09 (0.09)	0.01 (0.11)	0.16 (0.13)
0.8	7.78 (0.09)	0.18 (0.10)	**0.25** (0.12)	8.27 (0.19)	0.03 (0.21)	0.21 (0.22)
0.9	8.11 (0.07)	0.12 (0.10)	**0.28** (0.11)	8.83 (0.19)	-0.24 (0.23)	-0.14 (0.22)

	Naive	IPW	LIPW
		$\tau = 0.15$	
Intercept	2.764 (0.067)	1.995 (0.179)	1.948 (0.180)
Age	-0.304 (0.080)	-0.266 (0.191)	-0.261 (0.195)
Age^2	0.104 (0.048)	0.040 (0.102)	0.105 (0.103)
		$\tau = 0.25$	
Intercept	3.129 (0.096)	2.292 (0.128)	2.307 (0.127)
Age	-0.260 (0.087)	-0.292 (0.150)	-0.310 (0.147)
Age^2	0.123 (0.059)	0.092 (0.080)	0.096 (0.080)
		$\tau = 0.45$	
Intercept	3.896 (0.142)	2.900 (0.086)	2.859 (0.086)
Age	-0.209 (0.101)	-0.224 (0.083)	-0.270 (0.084)
Age^2	0.047 (0.070)	0.093 (0.056)	**0.120 (0.055)**
		$\tau = 0.50$	
Intercept	4.066 (0.127)	3.036 (0.090)	3.034 (0.092)
Age	-0.168 (0.097)	-0.245 (0.076)	-0.272 (0.078)
Age^2	0.047 (0.066)	0.100 (0.055)	**0.111 (0.056)**

Table 5: Coefficient estimation and bootstrap standard error from different methods for the Spain unemployment data set.

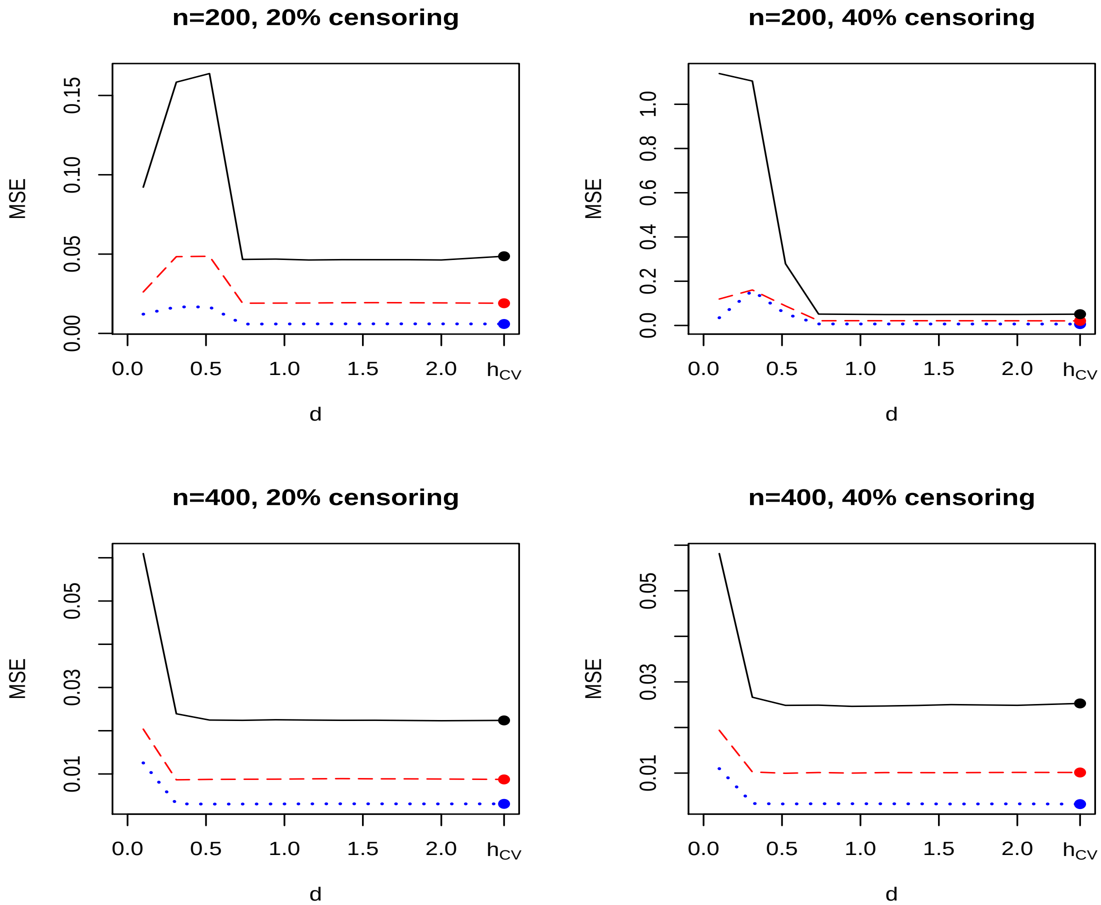

Figure 1: Mean squared error (MSE) of LIPW estimates with bandwidth $h_n = dn^{-0.3}$ in Case 2 at $\tau = 0.25$. For each curve, the solid point on the right end corresponds to the LIPW estimate with h_n selected by the five-fold cross validation (CV).

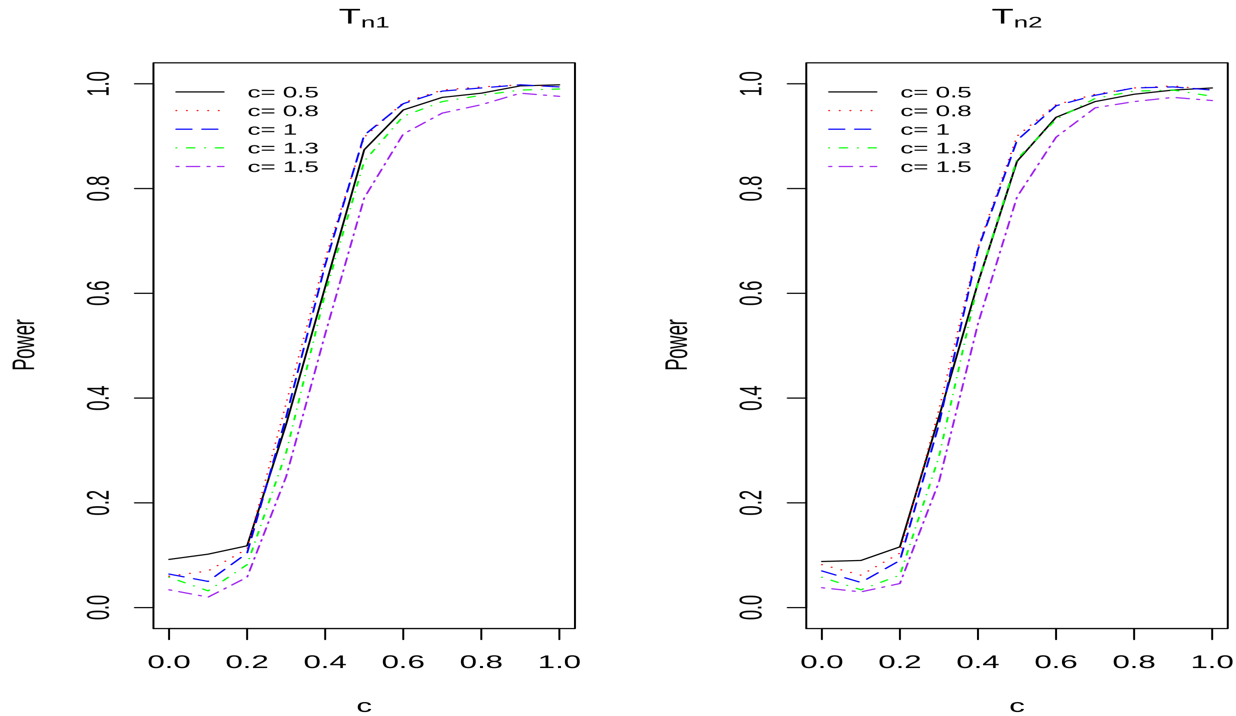

Figure 2: Simulated power curves of T_{n1} and T_{n2} with bandwidth $\lambda_n = cn^{-0.3}$ in Case 2 at $\tau = 0.5$.

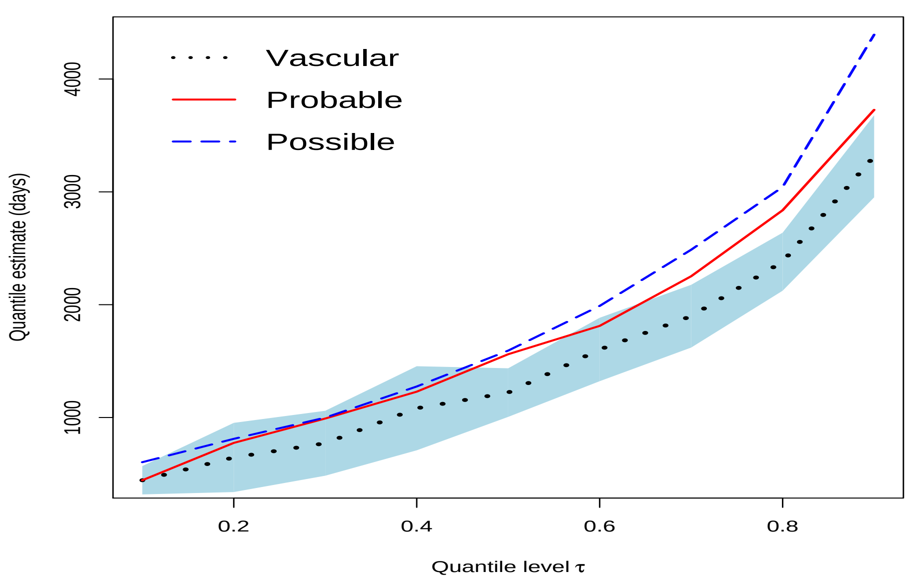

Figure 3: Estimated quantiles of population survival times for the three categories of dementia in the CSHA data set.

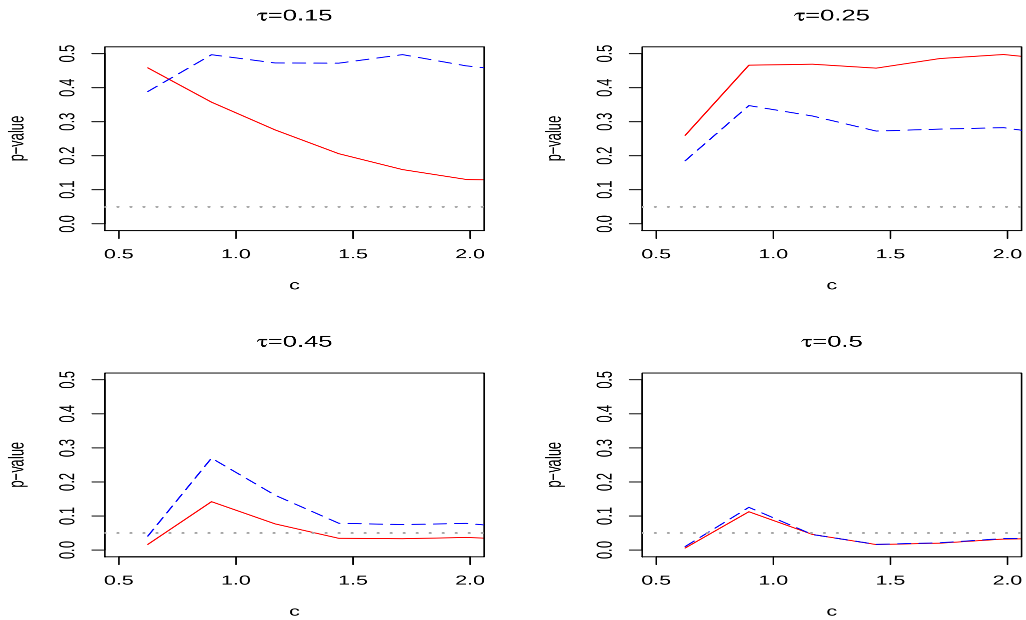

Figure 4: The p-values of the lack-of-fit test based on T_{n1} (solid curves) and T_{n2} statistics (dashed curves). The censoring distribution $G(\cdot)$ is estimated by using the global Kaplan-Meier estimates in T_{n1}, while by the local Kaplan-Meier estimates in T_{n2}. The dotted horizontal line corresponds to 0.05.

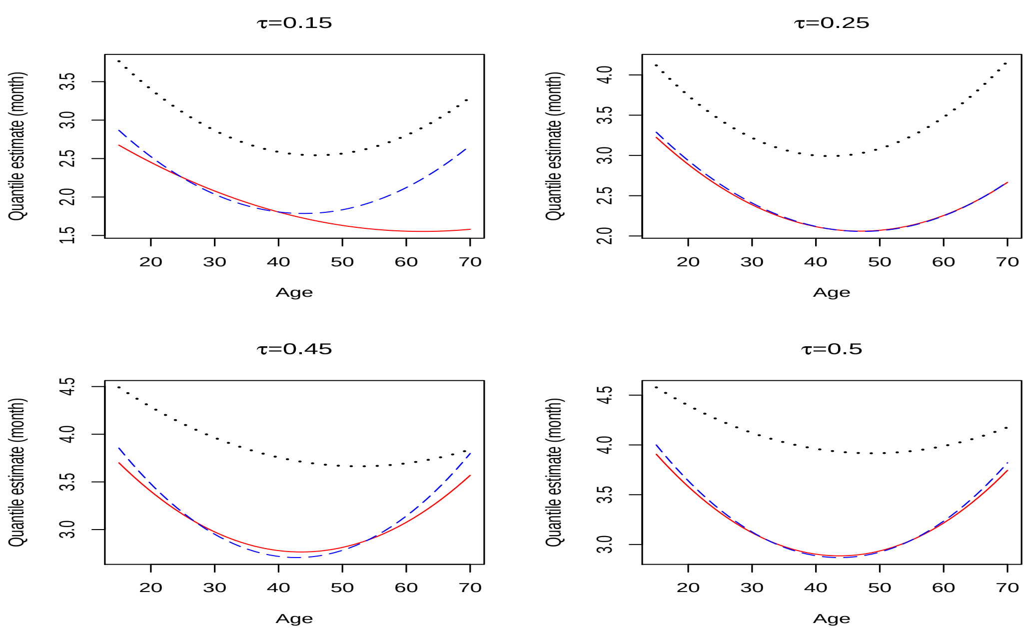

Figure 5: Estimated conditional quantiles of unemployment duration (in month) from different methods. The dotted, solid and dashed curves correspond to the Naive, IPW and LIPW methods, respectively.